

Db 1354 TATGCCCTTCTAGTGCCTATTTTGGCGGNGTCATGGTTCGACTTATGCTTGTTCATCAG 1413
Qy 466 ProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGly 485
Db 1414 CTTGTGTCTGTCTTCTCGGCAATGCGTTCTCCAAACTTCTCGAGCGGTATATT--- 1470
Qy 486 AspAspMetLysArgGluAsnProProValGluAspSerSerAspGluAspLysArg 505
Db 1471 -----GACCCCGTCATCCCGAAAGCGACGAGGAAGTCGCGAG 1509
Qy 506 AsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGlu 525
Db 1510 TCTCAGACGAGGTGTCTCAAGTCC---AAGCGCAAGAGATGCGCGCTGCCACGCC 1566
Qy 526 -----LysThrGluGluGlyLeu--- 531
Db 1567 AATAAGACGGGTCTCTTTACAGGTATTTTGGCGGCAAGTCTGTCTCCGCGCATTT 1626
Qy 532 GlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAla 551
Db 1627 GGTCTCGACACTCGATTGCTGTGGTTCCATTCTCTGTCTCTCTCTCTCTCTCTCT 1686
Qy 552 ValHisCysThrTrpValThrSerAsnAlaTyrSerProSerValValLeuAlaSer 571
Db 1687 CTTCACTGCACATGTGACTTCAACAGCGTATTCTTCGCTTCACTGTGCTACTTGCATCG 1746
Qy 572 TyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeu 591
Db 1747 CGAAACCGGATGTTAGCCAAATAATCATGATGATTCGAGAGCTTACTACTGGATT 1806
Qy 592 ArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIle 611
Db 1807 CGCAAAACACCCCGCAAGACAGCGTCATGCTGCTGCTGGGATTAACGGTACCAGATC 1866
Qy 612 AlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnSerHisIle 631
Db 1867 CTTGGTATGCTGATCGCCCCACCTTGTGTGATAACAATACCTGGAATAACCCACAT 1926
Qy 632 AlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr 651
Db 1927 GCCACAGTTGGTAAGCCATGGCTTCCACGAGATGTCGCATATCTTCTTGGAGAG 1986
Qy 652 LeuAspValAspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAsp 671
Db 1987 CATGATGTGATACATCTTCTGTGATCTTTGGGGCTTATTGGGTACTCTCGTCAGCAT 2046
Qy 672 IleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArg 691
Db 2047 ATCAACAAGTTTGTGGATGTTAGATCTCACAGGTGAATGGCTGACGAGGTGAG 2106
Qy 692 GluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThr 711
Db 2107 GAAGTCAACTACTTTTACTCAAGAGGGAGATGCTGTGATGACAGAGGCCACCCCTACT 2166
Qy 712 LeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAsp 731
Db 2167 ATGAAGACTCTCTCATGATCAAAATGCTTACTACCGTCTCCCGAGCTT----- 2217
Qy 732 PheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIle 751
Db 2218 TATGGTGGACACCCGCTCAAGACAGGGTTCGAGGCCAAATATCCCTCTAACAGTGT 2277
Qy 752 LysPheLysHisLeuGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLys 771
Db 2278 ACTCTTGATACTTTGACCAAGCGTTCAATCCGAAATTTGATCGTCAGGATCTACAAG 2337
Qy 772 ValLysAlaProAsp-----AsnArgGluThrLeuAspHisLys 784
Db 2338 GTCAAGAAGAGATCCCATTTGACGA-----GACCAACAAG 2373

RESULT 14

US-10-128-714-7139

; Sequence 7139, Application US/10128714

Publication No. US20030119013A1
GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wendi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Lemieux, Sebastien M
; APPLICANT: Eroshkin, Alexey M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7139
; LENGTH: 2232
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-7139

Alignment Scores:

Pred. No.:	9,46e-216	Length:	2232
Score:	2179.50	Matches:	416
Percent Similarity:	71.35%	Conservative:	117
Best Local Similarity:	55.69%	Mismatches:	177
Query Match:	49.52%	Indels:	37
DB:	15	Gaps:	8

US-10-028-384-2 (1-826) x US-10-128-714-7139 (1-2332)

Qy	69	GlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerArg	88
Db	55	CGCGAGCTTCTGGGATTTATTCTGTGTACCATTCGACGACGAGTTCGAGTCGA	114
Qy	89	LeuPheAlaValIleArgPheGluSerIleIleHisGluPheAspPheAspPheAsnTyr	108
Db	115	CTCTTCAGCTTATCGGCTTCGAGATATCATCAGAGTTGACCCGCGTTCACATTC	174
Qy	109	ArgSerThrHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGlu	128
Db	175	CGAGCAACAAATACCTAGTACAGAAATGGTTCTATAGCTTTTGGGATTTGGTATGAC	234
Qy	129	ArgAlaTrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMetIle	148
Db	235	GGAAATGGGATCTCTGGGACGTGTCACCGGTGGCAGCTTATATCCCGTCTCATGGTG	294
Qy	149	ThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAsp	168
Db	295	ACGAGCGCGTGTATCTACCATATCTTCGATCTCTTACTATCCCGTCGATATTCGCAAC	354
Qy	169	ValCysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeu	188
Db	355	ATTGCGCTCTACTGGCGCAGGATTCCTCGGCTGACTGCAATGGCAATGCTATGCTGTG	414
Qy	189	ThrArgGluLeuTrp---AsnGlnGlyAlaGlyLeuAlaAlaCysPheIleAlaIle	207
Db	415	ACATCCGAGATGCTCTCTCGCCATCTGAGGGTCTTCTTCGAGCAGCTTTCATGGGAATC	474
Qy	208	ValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIle	227
Db	475	GCCCTGTTTACATCTCCCGATCATGTTGTTGGAAGCTACGATAACGAAGCGATGCGATC	534

Qy 764 HisTrpLeuValArgIleTyrLys 771
 Db 2191 CATTTGGCTGGTCAGGATATACAG 2214

RESULT 12

US-10-417-375-97
 ; Sequence 97, Application US/10417375
 ; Publication No. US20040219528A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; APPLICANT: Marc Malandro
 ; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
 ; FILE REFERENCE: 529452001600
 ; CURRENT APPLICATION NUMBER: US/10/417,375
 ; CURRENT FILING DATE: 2003-04-15
 ; NUMBER OF SEQ ID NOS: 176
 ; SOFTWARE: Fast-Seq for Windows Version 4.0
 ; SEQ ID NO 97
 ; LENGTH: 5827
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-417-375-97

Alignment Scores:

Pred. No.: 1.8e-221 Length: 5827
 Score: 2241.00 Matches: 424
 Percent Similarity: 74.44% Conservative: 103
 Best Local Similarity: 59.89% Mismatches: 145
 Query Match: 50.92% Indels: 36
 DB: 18 Gaps: 7

US-10-028-384-2 (1-826) x US-10-417-375-97 (1-5827)

Qy 70 SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeu 89
 Db 181 ACATTTTGAAGCTTCTCATCTCTGTCATGGCTGCTGCTATATATCTCTCCACTGCTG 240
 Qy 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheSerTrpArg 109
 Db 241 TTTCGCTGCTGAGATTTGAAGTGTATCCATGAGTTTGTCCCGTACTTTAAATATCGG 300
 Qy 110 SerThrHisHisLeuAlaSerHisGlyPheTrpGluPheLeuAsnTrpPheAspGluArg 129
 Db 301 ACTACGAGGTCTCTGCTGAGGAGGGGTTTATAATTCATCACTGGTTTGCATGACCGA 360
 Qy 130 AlaTrpTyrProLeuGluValGlyValGlyThrValTyrProGlyLeuMetIleThr 149
 Db 361 GCCTGTACCCCTTTGGGAGCAATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCACC 420
 Qy 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
 Db 421 TCTGCTGCAATCTACCATCTACTCTCATTTTCCACATCACCACATTCGGAATGTC 480
 Qy 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 Db 481 TGTGTGTTCTGCGCCCT 540
 Qy 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
 Db 541 AAAGAGCTCAAGATGACGGGCTGGCTTCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 Qy 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 Db 601 GGATATATCTCCGATCTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 Qy 230 LeuGlnPheThrTyrTyrLeuTrpValIysSerValIysThrGlySerValPheTrpThr 249
 Db 661 ATGCTACTCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 720
 Qy 250 MetCysCysLeuSerTyrPheTrpMetValSerAlaTrpGlyGlyTyrValPheIle 269
 Db 721 GCTAAGTGTGCCCTTGTCTATTTCTACATGGTCTCGTCATSGGGAGGTTATGTTGTTCTG 780

Qy 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArg 289
 Db 781 ATCAACTTAATTCCTCTCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
 Qy 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
 Db 841 ATCATATGCTGCTACTGTACTGTTTACTGCTGGCAGCTATATCTTCTATGTCAGATCTCC 900
 Qy 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAla 329
 Db 901 TTTCGGGTTTCCAGCCCTGCTTTTCATCAGACACATGCGCAGCTTTGGGGTCTTTGGT 960
 Qy 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPhe 349
 Db 961 CTCTGCCAGATCCATGCTTTGTGATTAATCTGCGCAGCAAGTTGAATCCACAAATTT 1020
 Qy 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerVal 368
 Db 1021 GAAGTTCTTTTCCGGAGGCTCATCTCTCTG---GTAGGCTTTTGTCTCTCTCACCGTGGGA 1077
 Qy 369 IleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrp 388
 Db 1078 GCTCTCTCATGCTGACAGGAAATATCTCCCTGGAGCGGGGCTTTCTACTCGCTCTG 1137
 Qy 389 AspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnPro 408
 Db 1138 GATCCCTCTTATGCTAAGAACACATCCCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1197
 Qy 409 ThrThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGly 428
 Db 1198 ACAACCTGGTCTCATATTTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257
 Qy 429 LeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIle 448
 Db 1258 CTCTATTACTGCTTTAGCAACCTGCTGAGCGCGGATTTTATCATCATGATGCTG 1317
 Qy 449 SerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValVal 468
 Db 1318 ACCAGCATGACTTTTTCAGCTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1377
 Qy 469 CysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGlyAsp---Asp 487
 Db 1378 TGCAATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1437
 Qy 488 MetLysArgGluAsnProProValGluAspSerSerAspGluAspLysArgAsnGln 507
 Db 1438 ATAAAGTCTGCTCA----- 1449
 Qy 508 GlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLysThr 527
 Db 1450 -----GACAAAGAGACAGACAGACAGATTCACCTACCTATTAAAGAT 1497
 Qy 528 GluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuLeu 547
 Db 1498 GAA-----GTGGCAAGTGGGATGATGATGATGATGATGATGATGATGATGATGATG 1539
 Qy 548 MetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerVal 567
 Db 1540 ATCACTTACACCTTTCATTCACCTGGGTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1599
 Qy 568 ValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAla 587
 Db 1600 GTACTATCTCCCGTGTGGGATGCGAGTAGGATCATATTTGATGATCTCCGAGAGCA 1659
 Qy 588 TyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyr 607
 Db 1660 TATATTGCTGCTGCTCATATATCTCCAGAGGATGCGAGGATGCTGCTGCTGCTGCTGCTG 1719
 Qy 608 GlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn 627
 Db 1720 GGCTATCAGATTACAGCTATGCGAACCAGCAATTTTATGAGACATTAACATCATGGAAT 1779

[illegible]

QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
Db GGATATATCTCCGATCTGGTGGTGGTCTCTATGATAAAGAGGATTCCTATCTTTGC 631
QY 230 LeuGlnPheThrTyrLeuTrpValValSerValValThrGlySerValPheThr 249
Db ATGCTACTCCTACTACATGTGGATCAAGCAGTAAAGACTGGTTCATCTGTGGCA 691
QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
Db GCTAAGTGTGCGCTTGTCTATTTCTACATGCTCTGCTCATGGGAGGTTATGTTCTCG 751
QY 270 IleAsnLeuIleProLeuHlaValPheValLeuLeuMetGlnArgTyrSerLysArg 289
Db ATCACTTAACTCTCCAGCTCTGCTGATGCTCATGCGCCGTTCTCTACCGG 811
QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
Db ATCTATGGGCTACTGCTACTGTTTACTGCTGGTACTATCTTTCTAGGCAGATCTCC 871
QY 310 PheValGlyPheGlnProIleArgThrSerGluHlaMetAlaAlaIleValPheAla 329
Db TTTGGGTTTCCAGCTGCTCTTCTCATCAGACCATGCGAGGTTTGGGCTTTTGTGT 931
QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnIlePro 349
Db CTCTGCCAGATCCATGCTTGTGGATACCTGCGAGCAAGTTGAATCCACAAATTT 991
QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerVal --- 368
Db GAGTCTTTTCCGAGGCTCATCTCTG---GAGGCTTTGCTCTCTCACCCTGGGA 1048
QY 369 IleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrp 388
Db GCTCTCTCATGCTCACAGGAAATATCTCTGAGCGGGGCTTTCTACTCATCTGCTG 1108
QY 399 AspThrGlyTyrAlaLysIleHisIleProIleAlaSerValSerGluHisGlnPro 408
Db GATCCCTTTTATGTAAGAACACATCCCATCATCTCTGCTGCTGCTGCTGCTGCTG 1168
QY 409 ThrThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGly 428
Db ACAACCTGCTCATATATTGACCTGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1228
QY 429 LeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIle 448
Db CTCTATCTGCTTTAGCACTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1288
QY 449 SerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValVal 468
Db ACCAGCATGCTTTTCACTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1348
QY 469 CysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGlyAsp---Asp 487
Db AGCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1408
QY 488 MetLysArgGluAsnProProValGluAspSerSerAspGluAspLysArgAsnGln 507
Db ATAAGTCCGCA--- 1420
QY 508 GlyAsnLeuTyrAspLysAlaGlyValArgLysHisAlaThrGluGlnGluLysThr 527
Db -----GACAAAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1468
QY 528 GluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuLeu 547
Db -----GTGCAAGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1510
QY 548 MetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerVal 567
Db ----- 1511
Db ATCAGCTACCTTTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 1570
QY 568 ValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspPheArgGluAla 587

Db 1571 GTACTATCTCCGCTGCTGGGATGCGATGCGATAGATCATATTTGATGACTTCCGAGAGCA 1630
QY 588 TyrPheTrpLeuArgGlnAsnThrAspGluHisAlaGValMetSerTrpTrpAspTyr 607
Db TATATTTGGTCTGCTCATATCTCTCAGAGGATGCGAGGTCATGCTCTGGTGGGATAT 1690
QY 608 GlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnThrTrpAsn 627
Db 1691 GGTATCAGATTACAGCTATGCAACACCGAAATTTAGTGGCAATTAACACATGGAAT 1750
QY 628 AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLys 647
Db 1751 AATACCATATTTCTCAGTAGGCGAGCAATGGGCTCCACAGAGAAAACCTATGAG 1810
QY 648 IleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyr 667
Db 1811 ATCATGGAGGAGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1870
QY 668 SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAla-----Glu 683
Db 1871 TCCTCTGATGATCAACAAGTTTCTTTCGATGCTCCGATTTGGAGGAGCACAGATACA 1930
QY 684 GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
Db 1931 GGCACAT-----ATCAAGGAGATGACTATATATCTCAACTGGGAGTCCGT 1981
QY 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
Db 1982 GTGGACCGTGAAGTTCTCCAGTCTGCTCAACTGCTCATCTACAGATGTTGTACTAT 2041
QY 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArgThrArgAsn 743
Db 2042 CGCTTTGCGAGGTTTACAGAAAGCAAGCGCTCTCCAGGCTTTGACCGTCTCCGAAT 2101
QY 744 AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPheThrSerGlu 763
Db 2102 GCTGAGATTGGGAATAAGACTTTGAGCTTGATGCTCTCGAGGAAGGCTATACCAGAA 2161
QY 764 HistTrpLeuValArgIleTyrLysValLysAlaProAspAsnArg-GluThrLeuAspHis 783
Db 2162 CATTTGGTGTGAGGATATACAAAGGTAAGGACCTGGATAATCGAGGCTTGCAAGGACA 2221
QY 783 sLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrly 803
Db 2222 TAATGTCAGCTCCAGCTCTGATATCTTGGCACTGAGCACATCACATTTAGGACGTTGAA 2281
QY 803 s 803
Db 2282 G 2282
RESULT 11
US-10-417-375-99
; Sequence 99 Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 5404
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-99
Alignment Scores: 1.59e-221 Length: 5404
Pred. No.: 2241.00 Matches: 424
Score:

1229	CTCTATTACTGCTTTAGCAACCTGTCGTGATGCCCGGATTTTTTATCATCATGATGATGGTGTG	1288
449	SerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValVal	468
1289	ACACGATGATCTTTTCACCTGTAATGGTGGCGCTAAATGCTAGTGTGGCACCTGTTATG	1348
469	CysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGlyAsp---	487
1349	AGCATTTCTCTGGCATTTGGAGTCTCCAGGTCTGTCACACATACATAGGAATCTGGAC	1408
488	MetLysArgGluAsnProProValGluAspSerSerAspGluAspAspLysArgAsnGln	507
1409	ATTAAGTCGCCCA-----	1420
508	GlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLysThr	527
1421	-----GACAAAGAGCAAGAAGCAACAGGATTCACCTTACCTTATTAGATT	1468
528	GluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuLeu	547
1469	GAA-----GTGGCAAGTGGGATGATCTGTGTCATGGCTTTCCTTC	1510
548	MetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerVal	567
1511	ATCACCTACACCTTCATTCACCTGGTGACAGTCAGGCTTACTCTCTCCGTCCAIT	1570
568	ValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspPheArgGluAla	587
1571	GTACTATCTCCCGTGGTGGGATGGCAGTAGGATCATATTTGATGACTTCCGAGAAGCA	1630
588	TyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyr	607
1631	TATTATGGCTTCGTCAATACTCCAGAGGATCCGAAGTCATGCTCTGGTGGGATTAT	1690
608	GlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn	627
1691	GGCTATCAGATTACAGCTATGGCAAAACCGAACAATTTTAGTGACATAACACATGGAAT	1750
628	AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaIleTyrLys	647
1751	AATACCATATTTCTCGATAGGCAGCGCAATGGCGTCCACAGGAGAAAAGCCTATGAG	1810
648	IleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyr	667
1811	ATCATGAGGAGCTCGATGTCAGCTATGCTGTGTCATTTTGGAGGCCTCACTGGGTAT	1870
668	SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAla-----	683
1871	TCCTCTGATGATATCAACAAGTTCTTTGGATGTCGGATTGGAGGAGCACAGATACA	1930
684	GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg	703
1931	GGCAACAT-----ATCAAGAGGAATGACTATTATATCTCCAACTGGGGAGTTCCGT	1981
704	ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr	723
1982	GTGACCGTGAAGTTCTCCAGTCTGCTCAACTGGCTTCATGATACAGATGTGTACTAT	2041
724	ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn	743
2042	CGCTTTGGACAGGTTTACACAGAAGCCCAAGCGTCTCCAGGCTTTGACCGGTCCGAAAT	2101
744	AlaGluIleGlyAsnLysAspIleLysPheLysHisThrLeuGluAlaPheThrSerGlu	763
2102	GCTGAGATTGGGAATAAAGACTTTGAGCTTGATGCTCTGGAGGAAGCGCTATACCACGAA	2161
764	HisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArg-GluThrLeuAspHis	783
2162	CATTGGCTGGTCAGGATATACAGGTAAAGGACCTGGATATCGAGGGCTTGTCAAGGACA	2221
783	SlyProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLys	803
2222	TAAATGTCACGTCCAGCTCTGATATCTTCGACCTGAGCACATCATATTAGGAGCTTGAA	2281

Qy 803 s 803
 Db 2282 G 2282
 RESULT 10
 US-10-342-887-742
 ; Sequence 742, Application US/10342887
 ; Publication No. US20040058340A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dai, Hongyue
 ; APPLICANT: He, Yudong
 ; APPLICANT: Linsley, Peter S.
 ; APPLICANT: Mao, Mao
 ; APPLICANT: Roberts, Christopher J.
 ; APPLICANT: Van 't Veer, Laura Johanna
 ; APPLICANT: Van de Vijver, Marc J.
 ; APPLICANT: Bernards, Rene
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
 ; FILE REFERENCE: 9301-188-999
 ; CURRENT APPLICATION NUMBER: US/10/342,887
 ; CURRENT FILING DATE: 2003-01-15
 ; PRIOR APPLICATION NUMBER: 60/298,918
 ; PRIOR FILING DATE: 2001-06-18
 ; PRIOR APPLICATION NUMBER: 60/380,710
 ; PRIOR FILING DATE: 2002-05-14
 ; PRIOR APPLICATION NUMBER: 10/172,118
 ; PRIOR FILING DATE: 2002-06-14
 ; NUMBER OF SEQ ID NOS: 2699
 ; SEQ ID NO 742
 ; LENGTH: 2472
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-342-887-742
 Alignment Scores:
 Pred. No.: 5,168-223 Length: 2472
 Score: 2250.00 Matches: 432
 Percent Similarity: 73.01% Conservative: 109
 Best Local Similarity: 58.30% Mismatches: 163
 Query Match: 51.12% Indels: 37
 DB: 16 Gaps: 7
 US-10-342-887-742 (1-826) x US-10-342-887-742 (1-2472)

Alignment Scores:					
Pred. No.:	5.16e-223	Length:	2472		
Score:	2250.00	Matches:	432		
Percent Similarity:	73.01%	Conservative:	109		
Best Local Similarity:	58.30%	Mismatches:	163		
Query Match:	51.12%	Indels:	37		
DB:	16	Gaps:	7		
 US-10-028-384-2 (1-826) x US-10-342-887-742 (1-2472)					
QY	70	SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeu	89		
Db	152	ACACITTTTGAGCTTCTCATTCGTCAATGGCTGCTATTATCCCTTCCACTCGCTG	211		
QY	90	PheAlaValIleAtpPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArg	109		
Db	212	TTTGCTGCTCGAGAAATTTGAAGAAGTGTATCCAAGATTTGATCCGTACTTTAATATATCGG	271		
QY	110	SerThrHisHisLeuAlaSerHisGlyPheTyrgluPheLeuAsnTrpPheAspGluArg	129		
Db	272	ACTACCAGTTCTCGCTGAGGAGGGTTTTATAAATCCATAAGTGTTCATGACCGA	331		
QY	130	AlaAtPrpTrpProLeuGlyArgIleValGlyGlyThrValTyProGlyLeuMetIleThr	149		
Db	332	GCCTGTGATCCCTTTGGGACGAATCATTTGGAGAACAAATTTACCCAGTTTATGATCAC	391		
QY	150	AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal	169		
Db	392	TCTGCTGCAATCTACCATGTACTCCATTTTTTCCACATCACCATTCGGAATGTC	451		
QY	170	CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr	189		
Db	452	TGTGTGTTCTCGCCCCCTCTCTTCCTCCTTCCACCTCCATCGTACGTACCTCTTACC	511		
QY	190	ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro	209		
Db	512	AAGAGCTTAAGCATACAGGCTGGGCTTTCTGTGCTGCCATGATGTGTAGTAGTCT	571		

QY 668 SerGlyAspIleAsnLysPheLeuTyrMetValArgIleAla-----Glu 663
 DB 1871 TCCTCTGATGATATCAACAAGTTTCTTTGGATGGTCCGGATTGGAGGACACAGATACA 1930
 QY 684 GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
 DB 1931 GGCAGAAAT-----ATCAAGAGAAATGACTATTATATCTCAACTGGGGAGTTCCGT 1981
 QY 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
 DB 1982 GTGACCGTGAAGGTTCTCCAGTCTCTCACTGCTCACTGCTCACTGACAGAGTGTACTAT 2041
 QY 724 ArgPheGlyGluMetGlnLeuLeuAspPheArgThrProGlyPheAspArgThrArgAsn 743
 DB 2042 CGCTTTGGACAGGTTTACACAGAAGCCAGCGCTCTCCAGGCTTTCACCGTGCCGAAAT 2101
 QY 744 AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluIleLysPheThrSerGlu 763
 DB 2102 GCTGAGATTGGGAATAAGACTTTGAGCTTGATGTCTTGAGGAAGGCTATACACAGAA 2161
 QY 764 HistPheValArgIleTyrLysValLysAlaProAspAsnArg-GluThrLeuAspHi 783
 DB 2162 CATGGCTGTCTAGGATATACAAAGTAAGGACCTGGATAATCGAGGCTTGTCAGGACA 2221
 QY 783 sLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrLys 803
 DB 2222 TAAATGTCAGCTCAGCTCTGATATCTTCGCACTGAGCACATCACATTTAGGACGTTGAA 2281
 QY 803 s 803
 DB 2282 G 2282

RESULT 9

US-10-172-118-742
 ; Sequence 742, Application US/10172118
 ; Publication No. US20030224374A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dai, Hongyue
 ; APPLICANT: He, Yudong
 ; APPLICANT: Linsley, Peter
 ; APPLICANT: Mac, Mao
 ; APPLICANT: Roberts, Chris
 ; APPLICANT: Van 't Veer, Laura
 ; APPLICANT: Van de Vijver, Marc
 ; APPLICANT: Bernards, Rene
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
 ; FILE REFERENCE: 9301-195-999
 ; CURRENT APPLICATION NUMBER: US/10/172,118
 ; CURRENT FILING DATE: 2002-06-14
 ; PRIOR APPLICATION NUMBER: 60/380,770
 ; PRIOR FILING DATE: 2002-05-14
 ; NUMBER OF SEQ ID NOS: 2699
 ; SEQ ID NO 742
 ; LENGTH: 2472
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: NM_002219
 ; DATABASE ENTRY DATE: 2001-06-18
 US-10-172-118-742

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
5.16e-223	2250.00	2472	432
Percent Similarity:	73.01%	Conservative:	109
Best Local Similarity:	58.30%	Mismatches:	163
Query Match:	51.12%	Indels:	27
DB:	15	Gaps:	7

US-10-028-384-2 (1-826) x US-10-172-118-742 (1-2472)

QY 70 SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeu 89

DB 152 ACACCTTTGAAGCTTCTCATCTCTCAATGGCTGCTGTATTATATCTTCTCCACTCGTGTG 211
 QY 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArg 109
 DB 212 TTTGTGCTCTGAGATTTGAAAGTGTATCATCAGATTGTATCGTACTTTAATATATCGG 271
 QY 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTyrPheAspGluArg 129
 DB 272 ACTACACAGTTCTTGGCTGAGAGGGGTTTATTAATTTCCATAACTGGTTGTATGACCGA 331
 QY 130 AlaTrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThr 149
 DB 332 GCCTGGTACCTTTGGGACGAATCATTTGGAGAACAAATTTACCCAGGTTTAAATGATCACC 391
 QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
 DB 392 TCTGTGCTCAATCTACCATGTACTCCATTTTTCACATCCACATTCGACATTCGGAATGTC 451
 QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 DB 452 TGTGTGTTCTTGGCCCTTACC 511
 QY 190 ArgGlnLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
 DB 512 AAAGAGCTCAAGGATGAGGGGCTGGGCTTCTTCTGCTGCTGCCATGATGTCTGTAGTTCTCT 571
 QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 DB 572 GGATATATCTCCCATCTGTGGCTGGCTCTGTATGATTAATGAAGGATTCATCTTTTGC 631
 QY 230 LeuGlnPheThrTyrTyrLeuTrpValLysSerValLysThrGlySerValPheTrpThr 249
 DB 632 ATGCTACTACCTACTATCATGTGGATCAAGCAGTAAGACTGGTTCCATCTCTCTGGCA 691
 QY 250 MetCysCysLysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
 DB 692 GCTAAGTGTGCCCTTGGTTTATTTCTACATGCTCTGTCATGGGAGGTTAATGTTCTCTG 751
 QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArg 289
 DB 752 ATCAACTTAATCTCTCCACGCTCTGCTGATGTGTACAGCGCGTTTCTCTCACCAG 811
 QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
 DB 812 ATCTATGTGGCTACTGTACTGTTTACTGCTGGGACTATATCTTTCTAGGAGATCTCC 871
 QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAla 329
 DB 872 TTTGTGGGTTTCCAGCGCTGCTTTTCATCAGACACATGCGCAGGTTTGGGGTCTTTGGT 931
 QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPhe 349
 DB 932 CTCTGCCAGATCCATGCCCTTTGTGGATTACCTCGCAGCAAGTGAATCCACACAATTT 991
 QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyValPheLeuSerVal--- 368
 DB 992 GAAGTTCTTTTCCGGAGCGTCATCTCTCTG---GTAGGCTTTTCTCTCTCACCCTGGGA 1048
 QY 369 IleTyrLeuThrTyrThrGlyTyrIleAlaProTyrSerGlyArgPheTyrSerLeuTrp 388
 DB 1049 GCTCTCTCATGTGTGACAGGAAATATCTCTCGACGGGGGCTTTCTACTCACTGTCTG 1108
 QY 389 AspThrGlyTyrAlaLysIleHisIleProIleAlaSerValSerGluHisGlnPro 408
 DB 1109 GATCCCTCTTATGCTAAGAACAAATCCCATCATCTGCTGTCTGAGCATCAGCCC 1168
 QY 409 ThrThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGly 428
 DB 1169 ACAACCTGGTCTCATACTATTATTGACCTGACGCTCTCTCTCTCATGTTTCCAGTTGCG 1228
 QY 429 LeuTrpPheCysIleLysAsnIleAsnArgGluArgValPheValAlaLeuTyrAlaIle 448

Db 332 GCCTGGTACCCCTTTGGGACGAAATCATTTGGAGAACAAATTTACCCAGGTTTAAATGATCACCC 391
Qy 150 AlaGlyLeuIleHisTyrPheLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
Db 392 TCTGCTGCAATCTACCATGACTCTCAATTTTCCACATCACCATCCACATTCGGATGTC 451
Qy 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
Db 452 TGTGTGTCTCCGGCCCTCTCTCTCTCTCTCCATCCATCCATCCATCCATCCATCCATCC 511
Qy 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
Db 512 AAAGAGCTCAAGGATGAGGGGCTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 571
Qy 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
Db 572 GGAATATCTCCCGATCTGTGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 631
Qy 230 LeuGlnPheThrTyrTyrLeuTrpValIleSerValIleThrGlySerValPheThrThr 249
Db 632 ATGCTACTACCTACTACATGTGATCAAGGACGAGTAAAGACTGGTTCCATCTCTGGGCA 691
Qy 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIle 269
Db 692 GCTAAGTGTGCCCTTGCTTATTTCTCATGCTCTCGTCATGGGAGGTTATGTTCTCTG 751
Qy 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerIleArg 289
Db 752 ATCAACTTAAATCTCTCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 811
Qy 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuLeuSerMetGlnIlePro 309
Db 812 ATCATGTGGCTACTGACTGTTTACTGCTGGTACTATATCTTTCTAGGAGAGTCTCC 871
Qy 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
Db 872 TTGTGGGTTTCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 931
Qy 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrIleGlnLys 349
Db 932 CTCGCGCAGATCCATGCTTTGTGGATTACTCGCGAGCAAGTTGAATCCACAACTT 991
Qy 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerVal --- 368
Db 992 GAAGTCTCTTCCGAGCGTCTCTCTCTG---GTAGGCTTTCTCTCTCTCTCTCTCTCT 1048
Qy 369 IleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrp 388
Db 1049 GCTCTCTCTCATGCTGACAGGAAAAATATCTCCCTGGAGCGGGCTTTCTACTACTGCTG 1108
Qy 389 AspThrGlyTyrAlaLysIleHisIlePheIleAlaSerValSerGluHisGlnPro 408
Db 1109 GATCCCTCTTATGTAAGAACACATCCCATCATCTCTCTCTCTCTCTCTCTCTCTCTCT 1168
Qy 409 ThrThrTrpValSerPhePhePheAspLeuHisIleLeuValCysThrPheProAlaGly 428
Db 1169 ACAACCTGGTCTCATACTATTTTGGACTGTGAGCTCTCTCTCTCTCTCTCTCTCTCTCT 1228
Qy 429 LeuTrpPheCysIleLeuAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIle 448
Db 1229 CTCATTACTGCTTTAGCAACCTCTCTGATGCCCGGATTTTATCATCATGATGATGCTG 1288
Qy 449 SerAlaValTyrPheAlaGlyValMetValArgLeuLeuLeuThrLeuThrProValVal 468
Db 1289 ACCAGCATGTACTTTTCCGCTGTAATGGTGGCTCTATGCTAGTGTGGACCTGTATG 1348
Qy 469 CysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGlyAsp---Asp 487
Db 1349 AGCATCTCTCTGGCATTTGGAGTCTCCAGGTGCTGCCACATACATGAAGATCTGGAC 1408
Qy 488 MetLysArgGluAsnProProValGluAspSerSerAspGluAspAspIleArgAsnGln 507
...

Db 1409 ATAAGTGGCCCA----- 1420
Qy 508 GlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGlyThr 527
Db 1421 -----GACAAGAAGAGCAAGCAAGGATTCCACCTACCTATTAAAGATT 1468
Qy 528 GluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeu 547
Db 1469 GAA-----GTGGCAAGTGGGATGATCTGGTCTGATGGCTTCTCTCTCTCTCTCT 1510
Qy 548 MetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerVal 567
Db 1511 ATCACTTACACCTTTTCACTCAACCTGGGTGACCAAGTGGAGGCTACTCTCTCTCTCTCT 1570
Qy 568 ValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAla 587
Db 1571 GTACTATCTCCCGTGGTGGGATGCGATGAGTATATTTGATGACTTCCGGAAGCA 1630
Qy 588 TyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyr 607
Db 1631 TATTATTGGCTTCTCTATAATCTCTCAGAGGATGCGAAGGTCATGCTCTGGTGGATTAT 1690
Qy 608 GlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn 627
Db 1691 GGCTATCAGATTACAGCTATGGCAACCGAACAAATTTAGTGGACAAATACATCGGAAT 1750
Qy 628 AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLys 647
Db 1751 AATACCCATATTTCTCGAGTAGGGCAGGCAATGGGCTCCACAGAGGAAAAAGCCTATGAG 1810
Qy 648 IleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyr 667
Db 1811 ATCATGGAGGAGCTCGATGTCAGCTATGCTGGTCAATTTTGGAGGCTCTCTCTGGTAT 1870
Qy 668 SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAla-----Glu 683
Db 1871 TCCTCTGATGATATCAACAAGTTTCTTGGATGGTCCGGATTGGAGGGAGCACATACA 1930
Qy 684 GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
Db 1931 GGCAAACT-----ATCAAGGAGAAATGACTATATATCTCCAACTGGGAGGTCCGT 1981
Qy 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
Db 1982 GTGGACCGTCAAGGTTCTCCAGTGTCTCAACTGCCTCATGTACAGATGTGTACTAT 2041
Qy 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArgThrArgAsn 743
Db 2042 CGCTTTGGACGGTTTACAGANGCAAGCGTCTCCAGGCTTTGACCGGTTCGGAAT 2101
Qy 744 AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGlu 763
Db 2102 GCTGAGATTGGGAATAAAGACTTTGAGCTTGATGTCTCGAGGAGGCTATACCAAGAA 2161
Qy 764 HistripleuValArgIleTyrLysValLysAlaProAspAsnArg-GluThrLeuAspHis 783
Db 2162 CATTTGGTGTGAGGATATACAGGTAAGGACCTGGATATCGAGGCTTGTCAAGGACA 2221
Qy 783 sLysProArgValThrAsnIlePheProLysGlnIleTyrLeuSerLysLysThrThrly 803
Db 2222 TAAATGTCCAGTCCAGTCTGATATCTTCGACTGAGCACATCACATTTAGGACGTTGAA 2281
Qy 803 s 803
Db 2282 G 2282
RESULT 8
US-10-028-384-11
; Sequence 11, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPARIGENE INC.
; APPLICANT: PERREAULT, Claude


```
QY 439 GluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetVal 458
Db 1264 GTCGGATTTTATCATATGATGGTGGTACCGAGCATGACTTTTCAGCTGTAATGGTG 1323
QY 459 ArgLeuMetLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsn 478
Db 1324 CQTCTAATGCTGGTATTGGCACCTGTTATGTGCATCTTTCTGGCATGGTGTTCCTCCAG 1383
QY 479 ValPheGluHisTyrLeuGlyAsp---AspMetIleArgGluAsnProProValGluAsp 497
Db 1384 GTGCTGTCCACATATATGAAATATCGACATAAGTCGCCA-----1425
QY 498 SerSerAspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysVal 517
Db 1426 -----GACAAAGAAGCAAGAAG 1443
QY 518 ArgLysHisAlaThrGluGlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSer 537
Db 1444 CACAGGATTTCTACTTACCTTATTAAAGATGAG-----GTGGCGAGT 1485
QY 538 IleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
Db 1486 GGGATGATACTGGTTCATGCTTTTCTCATCACCTACACGTTTCATTCGACTTGGGTG 1545
QY 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThr 577
Db 1546 ACCAGTGAGGCTATTCTTCTCCCTCATGTACTGTCTGCTGCTGGTGGGATGCGAT 1605
QY 578 ArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGlu 597
Db 1606 AGGATCATTTTGTATGACTTCGAGAGCGTATTATTGGTCTCGTCCGTACATACTCCAGAG 1665
QY 598 HisAlaArgValMetSerTrpTrpAspTyrGlyGlnIleAlaGlyMetAlaAsnArg 617
Db 1666 GATGCAAGATCATGTATGGTGGATATTGGCTACCAATTAATCTGCAATGGCAATCGG 1725
QY 618 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 637
Db 1726 ACAATTTTAGTGGACATAACACATGGAATAATACCCATATTCTCGAGTAGGCGAGCA 1785
QY 638 MetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrVal 657
Db 1786 ATGGATCCACAGAGAAAGCCCTATGAATCATATGAGGAGCTTGATGTACGTATGTG 1845
QY 658 LeuValIlePheGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrp 677
Db 1846 CTTGTCTATTGGAGGCTTACTGGGTATTCTTCGGATCATATCAACAAGTTTCTTTGG 1905
QY 678 MetValArgIle---AlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPhe 696
Db 1906 ATGGTCCGATTTGAGGAGCAGACAGACAGGAGACACATTAAGGAGATGACTACTAT 1965
QY 697 ThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeu 716
Db 1966 ACTCTACTGGGGAATTCGTTGTATCGTGGAGGTTCTCCGGTGTCTCAACTGCCTT 2025
QY 717 MetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProPro 736
Db 2026 ATGTACAAAATGTGTACTACCGCTTTGGGAGCTTACACAGAACCCAGCGTCCACCA 2085
QY 737 GlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeu 756
Db 2086 GGTCTTGACCGTGTTCGAAATGCTGAGATTGGTAAATAAGACTTTGAGCTTGATGCTG 2145
QY 757 GluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAsp 776
Db 2146 GAGNAGCGTATACACAGACACTGGCTAGTCAGGATATACAGGTAAAGGACCTGGAT 2205
QY 777 AsnArg 778
Db 2206 AATCGA 2211
```

RESULT 6

```
US-10-417-375-95
; Sequence 95, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95
; LENGTH: 3046
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-95

Alignment Scores:
Pred. No.: 4,09e-224 Length: 3046
Score: 2262.00 Matches: 429
Percent Similarity: 74.41% Conservative: 103
Best Local Similarity: 60.00% Mismatches: 147
Query Match: 51.40% Indels: 35
DB: 18 Gaps: 7

US-10-028-384-2 (1-826) x US-10-417-375-95 (1-3046)

QY 70 SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerArgLeu 89
Db 181 ACATTTTGAAGCTTCTCATTTCTGTCATATGGTGTGTATTATCTCTTCCCATCGTCTG 240
QY 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 109
Db 241 TTTGCTGCTCGAGATTGAAGTGTATCCATGAGTTTATCCGACTTAAATATATCGG 300
QY 110 SerThrHisIleAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
Db 301 ACTACCAAGTTCCTGGCTGAGGAGGGTATTTATAATTCATATGTTTCATGACCGA 360
QY 130 AlaTrpTyrProGlnGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
Db 361 GCCTGGTACCTTTGGAGCAATCATTTGAGGAGCAATTTTACCAGGTTTAAATATCACC 420
QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
Db 421 TCTGCTCAATCTACCATGACTCCATCTTTTCCACATCATCCATCGACATTCGGAATGTC 480
QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
Db 481 TGTGTGTCTCGCCCTCTCTTCTCTCTCTCTCCACCATCGTACGATCCACCTTACC 540
QY 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
Db 541 AAAGAGCTCAAGATGACGAGGGCTGGGCTTCTGTCTGCCATGATTGCTGTAGTTCT 600
QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
Db 601 GGATATATCTCCGATCTGTGGCTGCTCTATGATGATGAAGGAGATTCGCACTTTTGC 660
QY 230 LeuGlnPheThrTyrLeuTrpValLysSerValLysThrGlySerValPheTrpThr 249
Db 661 ATGCTACTCACCTTACTATGATGATGATCAAGGAGTAAAGACTGTTCCATCTGTTGGCA 720
QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
Db 721 GCTAAGTGTCCTTCTCTTCTATGATGATGATGATGATGATGATGATGATGATGATG 780
QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerLysArg 289
Db 781 ATCAACTTAATCTCTCCACGCTCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840
QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuLeuSerMetGlnIlePro 309
```

Qy 658 LeuValIlePheGlyGlyValIleGlyTyrSerClyAspAspIleAsnLysPheLeuTyr 677
 Db 1846 CTTGTTCATTTTGGAGGCGCTTACTCGGTATCTTCGGATGATATCAACAGTTCTTTGG 1905
 Qy 678 MetValArgIle---AlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPhe 696
 Db 1906 ATGGTCCGGAATGGAGGAAGCACAGACAGGAAGACACATTAAGGAGAAATGACTACTAT 1965
 Qy 697 ThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrThrLeuLeuAsnCysLeu 716
 Db 1966 ACTCTACTGGGAATTCGGGTGTGATCGTGAGGTCTCCGGTGTCTGCTCAACGCTT 2025
 Qy 717 MetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProPro 736
 Db 2026 ATGTACAAAATGTGTACTACCGCTTTGGGCGAGTCTACACAGAAGCAAGCGTCCACCA 2085
 Qy 737 GlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeu 756
 Db 2086 GGCITTTGACCGTGTTCGAAATGCTGAGATTGGTAATAAAGACTTTTGAGCTTGATGCTCG 2145
 Qy 757 GluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAsp 776
 Db 2146 GAGGAAGCGTATACACAGAACACTGGCTAGTCAGGATATACAAAGCTAAAGCACCTGGAT 2205
 Qy 777 AsnArg 778
 Db 2206 AATCGA 2211

RESULT 5

US-10-028-384-9
 ; Sequence 9, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREAULT, Claude
 ; APPLICANT: MCBRIDE, Kevin
 ; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
 ; FILE REFERENCE: 5600-74
 ; CURRENT APPLICATION NUMBER: US/10/028,384
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 9
 ; LENGTH: 3094
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: NM_008408
 ; DATABASE ENTRY DATE: 2000-11-01
 ; RELEVANT RESIDUES: (1)..(3094)
 US-10-028-384-9

Alignment Scores:
 Pred. No.: 1,43e-224 Length: 3094
 Score: 2266.50 Matches: 435
 Percent Similarity: 70.73% Conservative: 104
 Best Local Similarity: 57.09% Mismatches: 152
 Query Match: 15.50% Indels: 71
 DB: 15 Gaps: 8

US-10-028-384-2 (1-826) x US-10-028-384-9 (1-3094)

Qy 61 GlyLeuSerGlnProAla---GlyTrpGln----- 69
 Db 7 GGGTTGAGTGGCGCTGACCGATGCGAGGGGAGCAGAGTGGTCTCTGAGGAGCATC 66
 Qy 69 ----- 69
 Db 67 CGTGAGGTATTGGAATATCATCATGTTGCCACCCATTGATGCAAGATGACTAAGCTTGA 126
 Qy 70 -----SerLeuLeuSerPheThrIleLeuPheLeu 79
 Db 127 TTTTGGGATTGCTCTATGAGACGAGCAGACACTCTTAAAGCTTCTCATCTCTGTCGATG 186

Qy 80 AlaTrpLeuAlaClyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIle 99
 Db 187 GCTGCTGTGTATCTTTTCTACTCGTCTTTTGTCTGTCTGAGATTTTGAAGTGTCAATC 246
 Qy 100 HisGluPheAspProTrpPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPhe 119
 Db 247 CATGAGTTTGTATCGTACTTTAATATATGCGACTACCCCGTTTCTGGCTGAGAGGGTTT 306
 Qy 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGly 139
 Db 307 TATAAATTCATAACTGGTTGATACCGGGTGTGGTACCCTTTGGCGCGAATCAITTGA 366
 Qy 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsnThr 159
 Db 367 GGAACAATTTACCAGGTTTAAATGATCACTTCTGCTGCAATCTACCATGTACTCATTTTC 426
 Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
 Db 427 TTCATATCACTATTCGACATTCGGAATGCTGTGTGTTCCTGGGCCCACTTTCTCTCT 486
 Qy 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValGlyLeu 199
 Db 487 TTCACCACCATCGTTACGTACCATCCCTTACCRAAGAGCTCAAGGATGCAGGAGCTGGGCTT 546
 Qy 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
 Db 547 CTTGCTGCTGCCATGATGCTGTCTGTCTGCTGGTATATTTCTCGATCTGTAGTGGCTCC 606
 Qy 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLys 239
 Db 607 TATGATATGAGGAATTCATCTTTTCATGCTGCTTACTTACTACATGTCGATCAAG 666
 Qy 240 SerValLysThrClySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMet 259
 Db 667 GCAGTGAAGACGTGTTCATCTATTGGGCTGCCAAGTGTGCCCTGTATTCTTACATG 726
 Qy 260 ValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
 Db 727 GTCTCTTCATGGGAGGCTATGTGTCTGTGATCAACTTGATTCCTCTACATGTCTGGTG 786
 Qy 280 LeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIle 299
 Db 787 CTAATGCTGACAGGCGGTTTCTCACCGGATCTACGTAGCTACTGTACTGTATTACTGC 846
 Qy 300 ValClyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
 Db 847 CTGGGACCATCTTTCTATGACAGATTTCTTTGTTGGTTTCCAGCGCGCTCTTTCATCA 906
 Qy 320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
 Db 907 GAAACATATGCGACGCTTTTGGAGTGTGTGCTCTCTCAGATCCCATGCTGTCTGATATAC 966
 Qy 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
 Db 967 CTGGGACGAGTTGATCCACAGCAATTCGAAGTCTTTTCCGAGGTTTATCTCCCTG 1026
 Qy 360 AlaAlaGlyAlaValPheLeuSerVal---IleTyrLeuThrTyrThrGlyTyrIleAla 378
 Db 1027 ---GTTGGGTTTGTCT 1083
 Qy 379 ProTrpSerGlyValArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIlePro 398
 Db 1084 CCTGGACAGGCGGTTTCTACTCTCTGCTGGATCCCTCTATGCTAAGATAACATTCCTC 1143
 Qy 399 IleIleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeu 418
 Db 1144 ATTATTGCACTGTTTCTGAGCAGCCAGCCCACTGGTCTTCTACTATTTTTCATCTA 1203
 Qy 419 HisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAsp 438
 Db 1204 CAGCTCCTTGTCTTCATGTTTCCAGTTGGCTCTATTACTTGTCTTTAGCAACCTGCTGAT 1263

Db CCGGTCCATATTCGTGACATCTGGGTCTCTGGCCGCGATCTTCAGTGGCTGACCTCC 457
Qy IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 202
Db ATCTCCACCTACCTGCTGACCAAGGAGCTGTGTCGCGGGCGCGCTCTTCGCGGCC 517
Qy CysPheAlaAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAsp 222
Db AGCTTCATCGCCATCGCGCTGGCTCATCAGTAGGTGGCTGGCTGGATCGTACCATAC 577
Qy GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLys 242
Db GAGGCAATTCATATTCGCGCTGCACTTCACTACTCTCTGCGGTGGCTCAGTGAAG 637
Qy ThrGlySerValPheThrTrpMetCysCysLeuSerTyrPheTyrMetValSerAla 262
Db ACTGATCCGCTGCTGGTGGCGCGGAGCCGCTTTGCTCTTCTACATGATGGTGGCGCC 697
Qy TrpGlyGlyTyrValPheIleAlaLeuIlePheProLeuHisValPheValLeuLeuLeu 282
Db TGGGTGGCTACGTGCTCATCATCACTGATACCCGCTGACGCTCTTCTGCTACTGCTCAT 757
Qy MetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
Db ATGGCAGTACTCGCGCGGTCTGCTGACCGCTACAGACCTTCTACATCTCTGGAGCTG 817
Qy IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
Db CTGTTCTCATGTCAGATCCCTCTGCGGATTCACACCGATACGACACGATGAAACATG 877
Qy AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgasp 342
Db GCTGGCTGGGAGTGTGTGCTCTTATGCGCGTGGCCACCTTGGCGCATTTGCAAGCC 937
Qy ArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGly 362
Db GTGCTGTGCGCAACGAGTTCGGAAGCTGTTTCATGTCGCGGATGCTGTGGCGGT 997
Qy AlaValPheLeuSerValIleTyrLeuTyrThrGlyTyrIleAlaProTrpSerGly 382
Db GCGGCTTTTGGCGCGTGGTGTGCTCACCATGCTGGCGTGTGGCGCGGTGGAGTGA 1057
Qy ArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSer 402
Db CGCTTCTACTGCTGTGGATCTGCTACGCTACGCAAGATCCATTCCTCATCTGATCC 1117
Qy ValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIleLeuVal 422
Db GTGTGGAGCATCAGCCACCATGTTGTTCTGCTTCTTGTATCTGCACATCTGGTG 1177
Qy CysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPhe 442
Db TGGCGCTTCCAGTGGGAGTGTGTACTGATCAAGCAGATCAACGACGAGCGGTTC 1237
Qy ValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 462
Db GTGGTGTGTACGCGCATCAGTGGCGTTACTTCTGCTGTGTGTGTGTGTGTGTGTGT 1297
Qy ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482
Db ACCCTCAGCGCGGT 1357
Qy TyrLeuGlyAspAsp---MetLysArgGluAsnProProValGluAspSerSerGlu 501
Db TTCCTGCAAGAGATTCGTCTAACGATGGGACACGCTATAGCCACGACCCACGAGTG 1417
Qy Asp-----AspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArg 518
Db GATGAAGCTGAGGATTCATTCAGAGAAGACGCTGTACGACAAAGCTGGCAAGCTG--- 1474
Qy LysHisAlaThrGluGlnGlyThrGluLys---GlyLeuGlyProAsnIleLysSer 537

Db AAGCATCGTACTAAGCATGATGCCAGCAGGATAGTGGCGTCACTCCAACTGAGAGT 1534
Qy IleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
Db ATTTGTTATTTGGCGGTCTTAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1594
Qy ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
Db ACCAGGAATGCTCTACTCCAGTCCCTCCATGTCTTGGTGTTCACACACAGTCAAGATGA 1654
Qy ThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAsp 596
Db TCCCGCAACATTTAGACGATTCAGAGAGGCTTACTACTGGCTTTCGCAAGAACACTGCC 1714
Qy GluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 616
Db GATGATGCTCGGTATGTCTTGGTGGATTAAGGATACAGATAGCGGGAATGGCAAC 1774
Qy ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 636
Db AGAACGACGCTAGTGGATAATAATACGTGAACATAGTACATAGAGCTGGTGGCAAG 1834
Qy AlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyr 656
Db GCATGCTCTCAACGAGGAGAGTCTTACGAAATTAATGATCTCTTGGCTGACTAC 1894
Qy ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 676
Db GTTTTGGTGTATCTTTGGCGGTGTGATGCTATTTCTGCGGATGATATCAACAGTTCTCT 1954
Qy TrpMetValArgIleAlaGluGlyGlnHisProLysAspIleArgGluSerAspTyrPhe 696
Db TGGATGCTCGAATGCTGAGGAGAGCATCCCAAGGACATTAAGAAAGCATTTACTTT 2014
Qy ThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeu 716
Db ACCGACGCGGTGAATTCAGGGTAGATCCGAAGTGTCTCCGCGCTGCTCAACTGCTT 2074
Qy MetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProPro 736
Db ATGTACAAATTAAGTACTACAGATTCGGAATTAAGTGTGGACTACAGAGTCCATCT 2134
Qy GlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeu 756
Db GGATATGATCGCACGACGTAACCGCTCATTTGGGAATAAGGACTTCGATCTGACCTGCT 2194
Qy GluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAsp 776
Db GAGAGGCTTACACACAGACACTGGCTTGTTCGATCTATAGGTGAGAGACCGCAT 2254
Qy -----AsnArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGln 794
Db GAGTTTCATAGACCATCACTGAAGACCAAGGAGAGACG-----ATTCTCCACGA 2305
Qy LysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuVal 814
Db AACTTCATTTCCAGAAAGAACTCTAAGCGTCCAGAGGCTACATACGAACCGACCGGTT 2365
Qy PheLysGlyGlyLysLysLysSerLysLysThr 825
Db GTTGTTTAAGGAAACGACCTTGAATAAACC 2398

RESULT 4

US-10-417-375-92
Sequence 92, Application US/10417375
Publication NO. US2004021952A1
GENERAL INFORMATION:
APPLICANT: David W. Morris
TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
FILE REFERENCE: 529452001600
CURRENT APPLICATION NUMBER: US/10/417,375
CURRENT FILING DATE: 2003-04-15

Db 1205 TGGAGTGGCAGGTTTATCTACATGGGATCGGATGATCAAAATACACATTCGATT 1264
Qy 400 lleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHis 419
Db 1265 ATTGCATCAGTGTGTGAACATCAGCCCTACGACATGGGTGTCTTCTTCTTTGATCTACAT 1324
Qy 420 lleLeuValCysThrPheProAlaGlyLeuTrpPheCysAlleLysAsnIleAsnAspGlu 439
Db 1325 ATCTCTGTATGATCTCTCCAGCAGGCTATGGTCTGCATCAAAATATCAACGATGA 1384
Qy 440 ArgValPheValAlaLeuTyAlaIleSerAlaValTyPheAlaGlyValMetValArg 459
Db 1385 AGAGTATTGTGCTGCTGTATGCGCATCAGTGTGTGTACTTTCGCGAGTGAATGGTGGG 1444
Qy 460 LeuMetLeuThrLeuProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
Db 1445 CTGATGCTGACTCTGACCCCGTGTCTGTCATGCTGTGCGCATCGCTTCTCCATGTT 1504
Qy 480 PheGluHisTyLeuGlyAspAspMetLysArgGluAsnProValGluAspSerSer 499
Db 1505 TTTGAGCACTATTGGGGGATGACATGAAAGGGAACCCACCTGTGGAGGACAGCAT 1564
Qy 500 AspGluAspAspLysArgAsnGlnGlyAsnLeuTyAspLysAlaGlyLysValArgLys 519
Db 1565 GATCAGGATGACAAAGAAACCCAGGAACTTGTATGACAAGCAGGTAAGTGAAGAG 1624
Qy 520 HisAlaThrGluGlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleVal 539
Db 1625 CATGTGACAGAGCAAGAGAACCTTGAAGAGGGCTTTGGGCCCCAACATCAAAAGCATGTG 1684
Qy 540 ThrMetLeuMetLeuMetLeuMetPheAlaValHisCysThrTrpValThrSer 559
Db 1685 ACCATGCTGATGCTCTGATGATGTTGCGGTCTCCACTGCAGTGGGTCAACAAGC 1744
Qy 560 AsnAlaTySerSerProSerValValLeuAlaSerTyAsnHisAspGlyThrArgAsn 579
Db 1745 AACGCTACTCAGTCCAGGTGTGGTCTTGCCTTCTACATCATGATGCTACAGGAAT 1804
Qy 580 lleLeuAspAspPheArgGluAlaTyPheTrpLeuArgGlnAsnThrAspGluHisAla 599
Db 1805 ATATTAGATGATTTAGAGAAGCGTACTTTTGGCTGAGACAAACACGAGTGAACAGCC 1864
Qy 600 ArgValMetSerTrpTrpAspTyGlyTyGlnIleAlaGlyMetAlaAsnArgThrThr 619
Db 1865 CGGGTCATGTCGTGTGGGACTACGGGTATCAGATTCGTGGCATGGCCCAACAGGACCAT 1924
Qy 620 LeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSer 639
Db 1925 CTGGTGGATAACAACACCTGGAACAACAGCCACATCGCACTGTGCGAAGCTATGTCT 1984
Qy 640 SerAsnGluThrAlaAlaTyLysIleMetArgThrLeuAspValAspTyValLeuVal 659
Db 1985 TCCATGAACGGCCGCTATAAATCATGAGGTCCCTTGTATGTCGATATGTTGGTT 2044
Qy 660 llePheGlyGlyValIleGlyTyTrpSerGlyAspAspIleAsnLysPheLeuTrpMetVal 679
Db 2045 ATTTCCGAGGAGTGAATGGCTATTCGGGAGCAGATATCAACAAGTTCTCTGGATGTC 2104
Qy 680 ArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyPheThrProGln 699
Db 2105 AGGATAGCTGAAGGGAGCATCCCAAGACATCCGGAAGGTGACATTTTCCACCCAGCAG 2164
Qy 700 GlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyLys 719
Db 2165 GGAGAGTTCCGAGTAGACAAAGCTGGTCTCTACTCTGTATAAATCGCTTATGATATAA 2224
Qy 720 MetSerTyTrpArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAsp 739
Db 2225 ATGTCATACTACAGATTGGAGAAATGAGTGTAGATTTTCCGACTCCCCAGGCTTTGAC 2284
Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAla 759

Db 2285 CGBACAGTAACTGCTGAGATTGGAATAAAGACATTAAATCAAGCATTTGAGGAAGCT 2344
Qy 760 PheThrSerGluHisTrpLeuValArgIleTyLysValLysAlaProAspAsnArgGlu 779
Db 2345 TTTACATCAGAGACCTGGCTTGTCAAGTATATAAAGTGAAGAACCTGTCACACAGGGAG 2404
Qy 780 ThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyLysSerLys 799
Db 2405 ACATAGGTCAACAACCTCGAGTCAACAACATCGTCCCAACACAGAATTTTGTCAAG 2464
Qy 800 LysThrThrLysArgLysArgGlyTyLysLysAsnLysLeuValPheLysLysGlyLys 819
Db 2465 AAGACTACTAAAGAGGAGCGTGGCTACGTAAATAAAGTAAAGTGTGTTTAAAGAAAGCAAG 2524
Qy 820 LysIleSerLysLysThrVal 826
Db 2525 AAGACCTCTAAGAAGACTGTT 2545

RESULT 3
US-10-028-384-7
; Sequence 7, Application US/10028384
; Publication No. US200301482B5A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Kevin
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5800-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7
; LENGTH: 2417
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AF132552
; DATABASE ENTRY DATE: 1999-04-27
; RELEVANT RESIDUES: (1)..(2417)
US-10-028-384-7

Alignment Scores:
Pred. No.: 7,3e-236 Length: 2417
Score: 2950.00 Matches: 557
Percent Similarity: 83.79% Conservative: 89
Best Local Similarity: 72.24% Mismatches: 113
Query Match: 67.03% Indels: 12
DB: 15 Gaps: 7

US-10-028-384-2 (1-826) x US-10-028-384-7 (1-2417)

Qy 63 SerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
Db 98 AGCAAGGTGGCTGGCTACAGAGCCCTAATCACTCGCCATCTCTGCTAATCGCTGGGTG 157
Qy 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
Db 158 GCGGATTTCTCTCGCTCTTCGCGCTATCCGTTTCGAGTCGATTCATCATGAGTTT 217
Qy 103 AspProTrpPheAsnTyArgSerThrHisHisLeuAlaSerHisGlyPheTyArgPhe 122
Db 218 GATCCGTGGTTCAACTACCGGCGCCACCGCTCATGTCAGTAATGGTGGTACAACTTC 277
Qy 123 LeuAsnTrpPheAspGluArgAlaTrpTyProLeuGlyArgIleValGlyGlyThrVal 142
Db 278 CTCAACTGGTTCGACGAGCGCATGGTATCGCTCGCAGGATTTGTGGCGGTACCGTC 337
Qy 143 TyProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle 162
Db 338 TATCCCGCGCTGATGATTACGTCCGCGGAATCCATTGCTGCTGCACGTACTCAACATA 397
Qy 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182

```

1981 TTTGGAGGGGTAATGGCTATCTGGTGATGATCAACAATTTCTCGATGGTGG 2040
Db
681 ILeAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
Qy
2041 ATAGTCGAAGAGAACATCCCAAGACATTCGGGAAGTGAATTTTACCCACAGGA 2100
Db
701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
Qy
2101 GAAATCCGTGTAGACAAAGCAGGATCCCTACATTTGTGAAATTCGCTTATGATAAAATG 2160
Db
721 SerTyrTyrArgPheGluMetGlnLeuAspPheArgThrProGlnGlyPheAspArg 740
Qy
2161 TCATCTACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCACAGGTTTACCGA 2220
Db
741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPhe 760
Qy
2221 ACAGTAAATGCTGAGATTTGAAATTAAGCAATTAATTAATTCAGCAATTTGGAAGACCTTT 2280
Db
761 ThrSerGluHisTyrPheValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
Qy
2281 ACATCAGAACACTGGCTTTGTAGATATATAAAGTAAAGCCCTGATTAACAGGAGACA 2340
Db
781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
Qy
2341 TTAGATCACAAACCTCGAGTCACCAACATTTTCCCAAAACAGAAATATTTGTCAAAGAAG 2400
Db
801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
Qy
2401 ACTACCAAGAGAGCGGTGCTACATTAATAAATAAGCTGTTTTTAAGAAGGCAAGAAA 2460
Db
821 IleSerLysLysThrVal 826
Qy
2461 ATATCTAAGAAGACTGTT 2478
Db

```

RESULT 2

```

US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)..(2469)
US-10-028-384-3

```

Alignment Scores:

```

Pred. No.: 0 Length: 2710
Score: 4263.50 Matches: 805
Percent Similarity: 79.70% Conservative: 3
Best Local Similarity: 97.34% Mismatches: 14
Query Match: 96.88% Indels: 5
DB: 15 Gaps: 2

```

US-10-028-384-2 (1-826) x US-10-028-384-3 (1-2710)

```

Qy 1 MetaLalGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTyr 20
Db 77 ATGGCGGAGCCCTCGGCCCCGGAGAGCAAGCAAGTCGCTCAACTCGTCCCCGGTGG 136

```

```

21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisHisGlyProGlyAlaGlnCys 40
Db
137 AGCGGCTCATGGCTCTGGGAAACAGCCGCCACGGGACCATGCGGCCGGAACCCAGAGC 196
Qy
41 AlaHisLysAlaAlaGlyAlaAlaProProLysProAla---ProAlaGlyLeuSer 59
Db
197 GCCTCCAGCGCGCGCGCG-----CCGAAGCGCGCGCGCGCGCGCGCGCTGTCC 244
Qy
60 GlyGlyLeuSerGlnProAlaGlyTyrGlnSerLeuSerPheThrIleLeuPheLeu 79
Db
245 GGCGGCTTGTCCAGCGCGCGCGGTGGCAGTCGTGTCTCTCTTCCATCTCTTCCCTG 304
Qy
80 AlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIle 99
Db
305 GCCTGCTGCCCGGCTTACGTCGCCCTCTTCCGCGTCATCCGCTTCGAGAGCATCATC 364
Qy
100 HisGluPheAspProTyrPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPhe 119
Db
365 CACGAGTTCGACCCGCTGGTTAACTATAGATCAACACATCATCTTGCATCTCATGATTC 424
Qy
120 TyrGluPheLeuAsnTyrPheAspGluArgAlaTyrTyrProLeuGlyArgIleValGly 139
Db
425 TATGAGTTTCTAAATTTGGTTTGTAAAGAGCATGGTACCACCTGGGAGAAATAGTGGT 484
Qy
140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleThrIleLeuAsnThr 159
Db
485 GGCACCGTTTACCAGGGTGTATGATAACAGCTGGCTTATTCTATTGGATTTTAAATACA 544
Qy
160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
Db
545 TTGAACATACAGGTTCACATAGAGATGTGTGTATTCCTTGCCACCACTTTTAGCGGC 604
Qy
180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTyrAsnGlnGlyAlaGlyLeu 199
Db
605 CTTTACATCCATATCTACGTTCTCTGCTAACTAGAGAACTGTGGAAACCAAGAGCAGACT 664
Qy
200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
Db
665 CTAGCTGCTTGTCTTANTGCTATCTACAGGGTACATATCTCGTACGTCGGGGATCC 724
Qy
220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValLys 239
Db
725 TTTGATAATGAAGGCAATGCCATTTTTCGGCTTCACTTCACTTCACTTATGGGTAAAG 784
Qy
240 SerValLysThrCysValPheThrMetCysCysCysLeuSerTyrPheTyrMet 259
Db
785 TCTGTGAAGACCGGGTCTGTGTCTGGACATGTGCTGCTGTCTGTCTGTCTGTCTGTCT 844
Qy
260 ValSerAlaTyrGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
Db
845 GTCTCTCGCGGGAGGTTATGTTTCACTCACTCACTCACTCACTCTCTCTCTCTCTCTCT 904
Qy
280 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIle 299
Db
905 TTCTCTGCTGATGCAGAGGACAGAGAGAGTCTATAGCATATAGCATATAGCATTTTACATT 964
Qy
300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
Db
965 GTGGGTTTAAATATATCCATGCAGATACCTTTTGTGGGATTTGAGCCCAATCAGAACAGC 1024
Qy
320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
Db
1025 GAGCAGATGGCAGCTGCAGGTGTCTTTCGGCTGTGTCGAGGTTTACGCTTTTTCAGTAT 1084
Qy
340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
Db
1085 CTGAGAGACCGGTTCACAAAACAGGAGTTCAGAGCCCTTTTCTTTTGGGTGTCTCACTA 1144
Qy
360 AlaAlaGlyValAlaPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
Db
1145 GCTGCGAGCGCTGTGTTTCTTGTGTCTATCTATCTGACATACAGGTTATATTGACCA 1204
Qy
380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399

```

OTHER INFORMATION:

US-10-028-384-1

Alignment Scores:

Pred. No.: 0
 Score: 4401.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 DB: 15

Length: 2481
 Matches: 826
 Conservative: 0
 Mismatches: 0
 Indels: 0
 Gaps: 0

US-10-028-384-2 (1-826) x US-10-028-384-1 (1-2481)

QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTrp 20
 DB 1 ATGGCGAGGCGCTCGGCGCGGAGCAAGCAAGTCGTCCTCAACTCGTCCCGGTGG 60
 QY 21 SerGlyLeuMetAlaLeuGlyVasnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 DB 61 AGTGGCTCATGGCCCTGGGANAACAGCCGCAAGCCAGCCAGCGGCGCGGCGCCAGTGC 120
 QY 41 AlaHisLysAlaAlaGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 DB 121 GCGACAAAGGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGGCGG 180
 QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 DB 181 GGGCTGTCGACCGCGCTGGGTGGCAGTCGCTCTCTCTCCATCCATCTCTCTCTCTG 240
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis 100
 DB 241 TGGCTTCGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCGCTTCGAAAGCATCATCC 300
 QY 101 GluPheAspProTrpPheAsnTrpArgSerThrHisLysLeuAlaSerHisGlyPheTrp 120
 DB 301 GAGTTCACCGCTGGTTCATTAATACATCAACATCACTCTGCTCATCTCATGGTTCAT 360
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGlyGly 140
 DB 361 GAATTTTAAATGGTTGTATGAAGAGCATGGTATCCATAGGAAGATAGTAGGTGTGT 420
 QY 141 ThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
 DB 421 ACTGTTTACCAGGTTGATGATAACCGCTGGCTTATTCATTTGGATTTTAAATACAT 480
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 DB 481 AACATAACTGTTCATCAAGAGACGATGTGTGTCTCTGCAACCACTTTTAGCGGCTT 540
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAlaGlyLeu 200
 DB 541 ACATCTATATCTACTTCTCTTACNAGAGAACTTTGGAAACNAGAGACAGGACTTTTA 600
 QY 201 AlalaCysPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySerPhe 220
 DB 601 GCTGCTTGTATTTATTTGTATTTGATACAGGCTACATATCTCGGTGAGTGTGGTCT 660
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpLeuTrpValLysSer 240
 DB 661 GATTAATGAAGCATTTGCTATTTTGGACATTTTGGACATTTTGGACATTTTGGACAT 720
 QY 241 ValLysThrGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMetVal 260
 DB 721 GTAAAAACTGGGTGAGTGTGTTTGGACAAATGTGCTGCTTATCTTATTTATATGTC 780
 QY 261 SerAlaTrpGlyGlyTrpValPheIleLeuAsnLeuIleProLeuHisValPheValLeu 280
 DB 781 TCTGCTGGGTGGTGTATTTATTAATCAATCTTATTTCCACTGCAATGATTTTGTGTG 840
 QY 281 LeuLeuMetGlnArgTrpSerLysArgValTrpIleAlaTrpSerThrPheTrpIleVal 300
 DB 841 TTACTGTGAGAGATACAGAAAGAGTCTACATAGCATATAGCATTTTCTACATTTGTG 900

QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 DB 901 GGTTTAAATATTAATGAGATACCTTTTGGGATCCAGCCAAATCAGAAAGTGA 960
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeu 340
 DB 961 CACATGCGAGCTGAGGTGTCTTTGATTCGTCGAGCTTATGCTTTCTTCAGTATCTG 1020
 QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 DB 1021 AGAGACGATTAAACAAACAGAGTTCAGACCCCTTTTGGGTGTATCACTAGCT 1080
 QY 361 AlaGlyAlaValPheLeuSerValIleTrpLeuThrTrpGlyTrpIleAlaProTrp 380
 DB 1081 GCAGGTGCTGTCTTCTTGTATGTCATCTATTGACTTATACAGGTACATGTCACATGG 1140
 QY 381 SerGlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaLysIleHisIleProIleIle 400
 DB 1141 AGTGGCAGGTTTATTTCACTTGGGATCTGGGTATGCAAAATACATCACTTCCATTTAT 1200
 QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePhePheAspLeuHisIle 420
 DB 1201 GCATCATGTCGTGAGCATCAACTAGACTTGGGTGTCTTTCTTCTTTCATCTACATATT 1260
 QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
 DB 1261 CTGTATGTACCTTCCAGCAGGCTTTGGTCTGCATCAAAATATCAACGATGAAGA 1320
 QY 441 ValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetValArgLeu 460
 DB 1321 GTATTTGTTGCTCTATATCAATCACTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 QY 461 MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
 DB 1381 ATGTTGACTTGTACTCCAGTCTGTGTATGCTGTCTGCTGCTGCTGCTGCTGCTGCT 1440
 QY 481 GluHisTrpLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSerAsp 500
 DB 1441 GAGCAGCTATTTGGGGATGACATGAAGAGGAAAAATCCACCTGTGGAGAGCAGCATGTAT 1500
 QY 501 GluAspAspLysArgAsnGlnGlyAsnLeuTrpAspLysAlaGlyLysValArgLysHis 520
 DB 1501 GAGGATGACAAAGAAACCAAGAAATTTGTATGATAGGAGCATTAAGTGAAGGAAACAT 1560
 QY 521 AlaThrGluGlnGlyThrGluGluGlyLeuGlyProAsnIleLysSerIleValThr 540
 DB 1561 GCAACTCAACAGGAAAAAATGAGAGGATAGGCTTATATATAATAAAGCATTTGTCACC 1620
 QY 541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
 DB 1621 ATGTTGATGCTGATGCTATTTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
 QY 561 AlaTrpSerSerProSerValValLeuAlaSerTrpAsnHisAspGlyThrArgAsnIle 580
 DB 1681 GCCTACTCTAGTCCAAAGTAGTCTGCTGCTCATCAATCATGATGGACAGGAAATATC 1740
 QY 581 LeuAspAspPheArgGluAlaTrpPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
 DB 1741 TTAGATGATTTTAGAAGACTTACTTTTGGCTTAGGCAAAATACAGATGAACATGCACGA 1800
 QY 601 ValMetSerTrpTrpAspTrpGlyTrpGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
 DB 1801 GTAACTGCTTGGTGGGATTTAGCTATCAGATAGCTGGAATGGCTAAATAGAACTACGTTG 1860
 QY 621 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
 DB 1861 GTGGATTAACACCTGGATTAACAGCCACATAGCACTGGTGGGAAAGCTATGCTTCT 1920
 QY 641 AsnGluThrAlaAlaTrpLysIleMetArgThrLeuAspValAspTrpValLeuValIle 660
 DB 1921 AATGAACACAGCAGCTTATTAATCATAGGACTCTAGATGATAGATTTATGTTTGGTATT 1980
 QY 661 PheGlyGlyValIleGlyTrpSerGlyAspAspIleAsnLysPheLeuTrpMetValArg

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 15, 2004, 06:02:06 ; Search time 1039.95 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-2
Perfect score: 4401
Sequence: 1 MAEPAPESKHSKSLNSP.....GVKNKLVFKGKISKXTV 826

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131523_8140/app.query.fasta_1.3740
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOGEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cgi -LIST=45 -DOCALIGN=200 -THR SCORE=ext -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODEL=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10028384@cgn_1_1_2577@runat_14122004_131523_8140
-NCPU=6 -ICPU=3 -NO_WMAP -LARGESUBSTR -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=130 -WARN TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2_6/ptodaa1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodaa1/pubpna/US06_NEW_PUB.seq*
- 3: /cgn2_6/ptodaa1/pubpna/US06_PUBCOMB.seq*
- 4: /cgn2_6/ptodaa1/pubpna/US07_NEW_PUB.seq*
- 5: /cgn2_6/ptodaa1/pubpna/US07_PUBCOMB.seq*
- 6: /cgn2_6/ptodaa1/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodaa1/pubpna/US08_NEW_PUB.seq*
- 8: /cgn2_6/ptodaa1/pubpna/US08_PUBCOMB.seq*
- 9: /cgn2_6/ptodaa1/pubpna/US09_PUBCOMB.seq*
- 10: /cgn2_6/ptodaa1/pubpna/US09_PUBCOMB.seq*
- 11: /cgn2_6/ptodaa1/pubpna/US09_PUBCOMB.seq*
- 12: /cgn2_6/ptodaa1/pubpna/US09_NEW_PUB.seq*
- 13: /cgn2_6/ptodaa1/pubpna/US10A_PUBCOMB.seq*
- 14: /cgn2_6/ptodaa1/pubpna/US10A_PUBCOMB.seq*
- 15: /cgn2_6/ptodaa1/pubpna/US10C_PUBCOMB.seq*
- 16: /cgn2_6/ptodaa1/pubpna/US10C_PUBCOMB.seq*
- 17: /cgn2_6/ptodaa1/pubpna/US10E_PUBCOMB.seq*
- 18: /cgn2_6/ptodaa1/pubpna/US10E_PUBCOMB.seq*
- 19: /cgn2_6/ptodaa1/pubpna/US10_NEW_PUB.seq*
- 20: /cgn2_6/ptodaa1/pubpna/US10_NEW_PUB.seq*
- 21: /cgn2_6/ptodaa1/pubpna/US60_NEW_PUB.seq*
- 22: /cgn2_6/ptodaa1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4401	100.0	2481	15	US-10-028-384-1	Sequence 1, Appli
2	4263.5	96.9	2710	15	US-10-028-384-3	Sequence 3, Appli
3	2950	67.0	2417	15	US-10-028-384-7	Sequence 7, Appli
4	2366.5	51.5	3093	18	US-10-417-375-92	Sequence 92, Appli
5	2366.5	51.5	3094	15	US-10-028-384-9	Sequence 9, Appli
6	2262	51.4	3046	18	US-10-417-375-95	Sequence 95, Appli
7	2250	51.1	2472	15	US-10-171-581-112	Sequence 112, App
8	2250	51.1	2472	15	US-10-028-384-11	Sequence 11, Appl
9	2250	51.1	2472	15	US-10-172-118-742	Sequence 742, App
10	2250	51.1	2472	16	US-10-342-887-742	Sequence 742, App
11	2241	50.9	5404	18	US-10-417-375-99	Sequence 99, Appl
12	2241	50.9	5404	18	US-10-417-375-97	Sequence 97, Appl
13	2181.5	49.6	2466	16	US-10-320-797-2305	Sequence 2305, Ap
14	2179.5	49.5	2466	16	US-10-128-714-7139	Sequence 7139, Ap
15	2164	49.2	2256	15	US-10-032-585-6323	Sequence 6323, Ap
16	2120.5	48.2	2779	17	US-10-437-963-99904	Sequence 99904, A
17	2118.5	48.1	2839	18	US-10-425-115-150745	Sequence 150745,
18	2117	48.1	2681	16	US-10-425-114-14408	Sequence 14408, A
19	2023	46.0	2603	15	US-10-128-714-6139	Sequence 6139, Ap
20	2023	46.0	2603	15	US-10-128-714-139	Sequence 139, App
21	2023	46.0	4603	15	US-10-128-714-5139	Sequence 5139, Ap
22	2021.5	45.9	2157	9	US-09-801-368-387	Sequence 387, App
23	2021.5	45.9	2157	18	US-10-793-639-318	Sequence 318, App
24	2021.5	45.9	2733	15	US-10-028-384-5	Sequence 5, Appli
25	1958	44.5	1828	10	US-09-945-527-62	Sequence 62, Appl
26	1941.5	44.1	4738	16	US-10-320-797-305	Sequence 305, Appl
27	1939.5	44.1	2882	16	US-10-320-797-1305	Sequence 1305, Ap
28	1897	43.1	1848	15	US-10-128-714-2139	Sequence 2139, Ap
29	1896.5	43.1	1848	15	US-08-974-879-133	Sequence 133, App
30	1896.5	43.1	1543	10	US-09-305-736-133	Sequence 133, App
31	1896.5	43.1	1543	10	US-09-818-683-133	Sequence 133, App
32	1896.5	43.1	1543	11	US-09-818-683-133	Sequence 133, App
33	1896.5	43.1	1543	16	US-10-621-401-133	Sequence 133, App
34	1894.5	43.0	1209	15	US-10-106-698-330	Sequence 330, App
35	1894.5	43.0	1209	16	US-10-264-237-412	Sequence 412, App
36	1882.5	42.8	3141	18	US-10-425-115-130787	Sequence 130787
37	1855	42.1	1369	15	US-10-128-714-1139	Sequence 1139, Ap
38	1851.5	42.1	3397	16	US-10-424-599-111541	Sequence 111541,
39	1610.5	36.6	2244	17	US-10-437-963-48342	Sequence 48342, A
40	1581	35.9	2660	16	US-10-264-049-630	Sequence 630, App
41	1552.5	35.3	1114	16	US-10-296-115-629	Sequence 629, App
42	1388.5	29.3	1728	16	US-10-424-599-122476	Sequence 122476,
43	975.5	22.2	1094	17	US-10-437-963-99902	Sequence 99902, A
44	871	19.8	500	9	US-09-998-538-1643	Sequence 1643, Ap
45	763	17.8	485	10	US-09-918-955-11283	Sequence 11283, A

ALIGNMENTS

RESULT 1
US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: FERRELL, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)

; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3090
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-3090

Alignment Scores:

Pred. No.:	2,03e-50	Length:	594
Score:	544.00	Matches:	102
Percent Similarity:	78.31%	Conservative:	28
Best Local Similarity:	61.45%	Mismatches:	32
Query Match:	12.36%	Indels:	4
DB:	4	Gaps:	3

US-10-028-384-2 (1-826) x US-09-248-796A-3090 (1-594)

QY	76	IleLeuPheLeuAlaTrpLeuAlaGly-----PheSerSerArgLeuPheAlaValIle	93
DB	100	ATTATATTATATCG--ATAGCAGGTGCAGCTATTCTCTCGTTTATTTCCGTGATT	156
QY	94	ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgSerThrHisHis	113
DB	157	CGATTGGAAGATTATATTCATGAAATCGATCCCTGGTTCATTTCCGACACCAATAT	216
QY	114	LeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro	133
DB	217	TTAGTCACTCATTCCTTTATGAATTTTGAATTTGGTTTGATGATAGAATTGGTACCCA	276
QY	134	LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIle	153
DB	277	TGGGAAGAGTCACTGGTGTTATATCCCGTTTATGGTGACTTCAGTGCCCAT	336
QY	154	---HisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspValCysValPhe	172
DB	337	TGGCATATTTCACGTGATTGGTTTGCCTTACCGTTTGATATTATAAATATTGTGTTTA	396
QY	173	LeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeu	192
DB	397	TTAGCACCATTTTCTCGGATTAACTGCAATTTGACTTATTTTGTACTAAAGAAATG	456
QY	193	TrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIle	212
DB	457	AACGATTTCTAGTCACGATTATTGGCAGCTATATTATGGGATTGCCCCAGGTTATATT	516
QY	213	SerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPhe	232
DB	517	TCAAGATCAGGGCTGGTCTTATGATPATGAAGCAATTGCCAATTACTTTTAAATGGCA	576
QY	233	ThrTyrTyrLeuTrpVal	238
DB	577	ACATTTTATTTCTGGATT	594

Search completed: December 15, 2004, 11:58:29
Job time : 200.664 secs

QY 574 HisAspGlyThrArg 578
 DB 21 CAAGATGGATCCCGC 7

RESULT 13

US-09-270-767-16600/c
 ; Sequence 16600, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 16600
 ; LENGTH: 560
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-16600

Alignment Scores:
 Pred. No.: 3,68e-56 Length: 560
 Score: 595.50 Matches: 119
 Percent Similarity: 78.38% Conservative: 26
 Best Local Similarity: 64.32% Mismatches: 33
 Query Match: 13.53% Indels: 7
 DB: 4 Gaps: 5

US-10-028-384-2 (1-826) x US-09-270-767-16600 (1-560)

QY 400 IleAlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHis 419
 DB 558 ATTGCATCCGTCGGAGCATGACCCACCATGTTGTTCTGTTCTTTCATCTGCAC 499
 QY 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysAlaIleValAsnIleAspGlu 439
 DB 498 ATCATGTGTCGGCTTCCAGTGGGAGTGTGTACTGTCATCAAGCAGATCAACACGAG 439
 QY 440 ArgValPheValAlaLeuValSerAlaValThrPheAlaGlyValMetValArg 459
 DB 438 CGCGTTTTCGTGGTGTGTACGCCATCAGTGGGTTTACTTCGTTGTTGTGTGCGT 379
 QY 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
 DB 378 TTGATGTGACCTTCAGCGCGGTGGTGTGCATGCTGGCGGAGTGGCCCTTTTCGGGACTG 319
 QY 480 PheGluHisValLeuGlyAspAsp---MetLysArgGluAsnProProValGluAspSer 498
 DB 318 TTGCATGTGTCCTGCAAGAGGATTCGCTTAACCGAATGGGCACAGCCATAGCCAGCC 259
 QY 499 SerAspGluAsp-----AspLysArgAsnGlnGlyAsnLeuTyAspLysAlaGly 515
 DB 258 ACCGAAGTGAAGTGAAGTGAAGTTCATTGAGAGAGAGACCGCTGTACACAAAGCTGGC 199
 QY 516 LysValArgLysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsn 534
 DB 198 AAGCTG---AAGCATGCTACTAAGCATGATGCCAGCAGGATGCTGGCTCAGCTCCAAAC 142
 QY 535 IleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCys 554
 DB 141 CTGAAGAGATATGTTATTTTGGCGGTCTTAATGCTGTTGATGATGTTGCTGCTCCACTGC 82
 QY 555 ThrTrpValThrSerAsnAlaValSerSerProSerValValLeuAlaSerValAsn--- 573
 DB 81 ACGTGGGTGACCAAGCATGCTACTCCAGTCCCTCCATGTTGCTTTCACCAACAGT 22
 QY 574 HisAspGlyThrArg 578
 DB 21 CAAGATGGATCCCGC 7

RESULT 14

US-09-270-767-27262
 ; Sequence 27262, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 27262
 ; LENGTH: 503
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-27262

Alignment Scores:
 Pred. No.: 1.53e-51 Length: 503
 Score: 553.00 Matches: 105
 Percent Similarity: 75.30% Conservative: 20
 Best Local Similarity: 63.25% Mismatches: 41
 Query Match: 12.57% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-27262 (1-503)

QY 210 GlyTyIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 DB 2 GGGTATATCTTCGATCCGTCGGCGGATCGTACGACATGAGCGATCGCCATTTCTGC 61
 QY 230 LeuGlnPheThrTyTyIleuTrpValLysSerValLysThrGlySerValPheThr 249
 DB 62 ATGCTCTTCCATCTACTTTCGATCAAGCGGTAAAGACGGCAGCATCTTTTGTGTCG 121
 QY 250 MetCysCysLeuSerTyPheTyMetValSerAlaTrpGlyTyValPheIle 269
 DB 122 GCTATGTCGGCATTTGGCTACTTCTATATGTTCTCTCGTGGGCTGATGCTTCTG 181
 QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTySerLysArg 289
 DB 192 ATTAACCTAATCCCGCTGCAGCTGCTGGCGGTGATGATCACCGGACGTTTCTCGCAGG 241
 QY 290 ValTyIleAlaTySerThrPheTyIleValGlyLeuLeuLeuSerMetGlnIlePro 309
 DB 242 AICTATACATAGCATACAGACGCTATACGTCGGCACCATTCTGTCGATGCAGATCTCG 301
 QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
 DB 302 TTGTGGGATTCGAACCCCATCCAGAGCTCCGAACACATGCTGGCAGTGGGAACCTTTGGC 361
 QY 330 LeuLeuGlnAlaTyAlaPheLeuGlnTyIleuArgPheArgLeuThrLysGlnGluPhe 349
 DB 362 CTGTGCGATTCAGCTTTTCGCACTATCTGGCTCGGCATTCAGGATGCATCTTC 421
 QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValle 369
 DB 422 GATCTCTCTTCAAGACAGTGGTTTCCAGTGTGTTTACTGTGTGTGTTCTGCTGGGTACC 481
 QY 370 TyrLeuThrTyThrGly 375
 DB 482 CTGCTCAGCTTACCGGG 499

RESULT 15

US-09-248-796A-3090
 ; Sequence 3090, Application US/09248796A
 ; Patent No. 6747137
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstock et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A

RESULT 11

US-09-328-111-332
 ; Sequence 332, Application US/09328111
 ; Patent No. 6262333
 ; GENERAL INFORMATION:
 ; APPLICANT: Endege, Wilson O.
 ; APPLICANT: Steinmann, Kathleen E.
 ; APPLICANT: Astle, Jon H.
 ; APPLICANT: Burgess, Christopher C.
 ; APPLICANT: Bushnell, Steven E.
 ; APPLICANT: Carroll III, Eddie
 ; APPLICANT: Catino, Theodore J.
 ; APPLICANT: Derti, Adnan
 ; APPLICANT: Ford, Donna M.
 ; APPLICANT: Lewis, Marcia E.
 ; APPLICANT: Monahan, John E.
 ; APPLICANT: Schlegel, Robert
 ; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
 ; TITLE OF INVENTION: PRODUCTS
 ; FILE REFERENCE: CCD-257 (US)
 ; CURRENT APPLICATION NUMBER: US/09/328.111
 ; CURRENT FILING DATE: 1999-06-08
 ; EARLIER APPLICATION NUMBER: US 60/088,801
 ; EARLIER FILING DATE: 1998-06-10
 ; NUMBER OF SEQ ID NOS: 850
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 332
 ; LENGTH: 616
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1) - (616)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-328-111-332

Alignment Scores:

Pred. No.: 4,33e-68 Length: 616
 Score: 704.00 Matches: 148
 Percent Similarity: 86.29% Conservative: 3
 Best Local Similarity: 84.57% Mismatches: 17
 Query Match: 16.00% Indels: 8
 DB: 3 Gaps: 2

US-10-028-384-2 (1-826) x US-09-328-111-332 (1-616)

Qy 208 ValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIle 227
 Db 5 GTACAGGCTACATATCTCGTCACTAGCTGGATCTTTGATATGAGGCAATTCCTATT 64
 Qy 228 PheAlaLeuGlnPheThrTyrTyrLeuTyrValIlySerValIlySerThrGlySerValPhe 247
 Db 65 TTTCACCTTCAGTTACATATCTTTATGGTAAATCTGTAAAACTGGGTCACTTTT 124
 Qy 248 TrpThrMetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrVal 267
 Db 125 TGGCAATGTCTCTCTCTATCTATCTATCTCTCTCTCTCTCTCTCTCTCTCTCT 184
 Qy 268 PheIleIleAnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSer 287
 Db 185 TTATCATCAATCTTTATTCACATGATATTTGTGTGTGTGTGTGTGTGTGTGTGTGT 244
 Qy 288 LysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGln 307
 Db 245 AAAAGAGTCTACATAGCATATAGCATTTCTACATTTGTGGTTTATATATATATATG 304
 Qy 308 IleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyVal 327
 Db 305 ATACCTTTTGGGATTCCAGCCCAATCAGAACAAAGTGAACACATGGCAGCTTGCAGGTG 364
 Qy 328 PheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrIleuArgAspArgLeuThrIysGln 347
 Db 365 TTTCATTTGCTGCAAGCTTAANCCTTTCTTCAGTATCTGAGAC-CGATTACCAACCA 423

Qy 348 GluPheClnThrLeuPhePheLeuGlyValSerLeu-AlaAlaGlyAlaValPheLeuSe 367
 Db 424 GAGTTCAGACACCTTTTCNTTTTGGGGATACACTACTCAGNGCTGGTCTCTGANGCATAT 483
 Qy 367 rValIleTyrLeuThrTyrThrGlyTyrIleAlaPro---Trp 380
 Db 484 TGNATC-----NGGTACATTCCTCCCTCGATGG 511

RESULT 12

US-09-270-767-1318/c
 ; Sequence 1318, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Honburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1318
 ; LENGTH: 560
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-1318

Alignment Scores:

Pred. No.: 3,68e-56 Length: 560
 Score: 595.50 Matches: 119
 Percent Similarity: 78.38% Conservative: 26
 Best Local Similarity: 64.32% Mismatches: 33
 Query Match: 13.53% Indels: 7
 DB: 4 Gaps: 5

US-10-028-384-2 (1-826) x US-09-270-767-1318 (1-560)

Qy 400 IleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHis 419
 Db 558 ATTGCATCCGTGCGGAGCATCAGCCACCATCTGGTCTCTCTCTCTCTCTCTCTCT 499
 Qy 420 IleLeuValCysThrPheProAlaGlyLeuTyrPheCysIleLeuAsnIleAsnAspGlu 439
 Db 498 ATCATGTTGGCCCTTCCCATGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 439
 Qy 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
 Db 438 CGCGTTTTCGTGGTGTGTGTACCCCATCAGTCGGGTTTACTTCGTGTGTGTGTGTGT 379
 Qy 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
 Db 378 TTGATGTTGACCTCAGCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 319
 Qy 480 PheGluHisTyrLeuGlyAspAsp---MetLysArgGluAsnProProValGluAspSer 498
 Db 318 TTGATGTTGTTCTCTGCAAGAGGATTCGTCTAAGCAATGGGCACAGCCATAAGCGCAC 259
 Qy 499 SerAspGluasp-----AspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGly 515
 Db 258 ACCGAAGTGTGAAGCTGAGGATTCCTCAATGAGAGAAGACAGCTGTACGACAGGCTG 199
 Qy 516 LysValArgLysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsn 534
 Db 198 AAGCTG---AAGCATGCTACTAAGCATGATGCCAGCAGGATCTGGCGTCAGCTCCAA 142
 Qy 535 IleLysSerIleValThrMetLeuMetLeuMetMetMetMetMetMetMetMetMetMet 554
 Db 141 CTGAAAGATATGTTATTTTGGCCGTTCTTAATGCTGTGTGTGTGTGTGTGTGTGTGTGT 82
 Qy 555 ThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn--- 573
 Db 81 ACGTGGGTGACCAAGCAATGCCTACTCCAGTCCCTCCCTCTCTCTCTCTCTCTCTCTCT 22

Db 205 ATACGAACCCGCGGTGGTGTGTTAGGGAAACGAACCTTGAATAAACCC 155

RESULT 9

US-09-270-767-15440/c

; Sequence 15440, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 15440

; LENGTH: 1660

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

US-09-270-767-15440

Alignment Scores:
Pred. No.: 1,14e-78 Length: 1660
Score: 806.50 Matches: 153
Percent Similarity: 82.03% Conservative: 25
Best Local Similarity: 70.51% Mismatches: 34
Query Match: 18.33% Indels: 5
DB: 4 Gaps: 2

US-10-028-384-2 (1-826) x US-09-270-767-15440 (1-1660)

```
Qy 611 IleAlaGlyMetAlaAsnArgThrLeuValAspAsnAsnThrTrpAsnAsnSerHis 630
Db 796 GTTCGGGAGATGCAACAGACACGCTAGTGGATTAATACGTGGAACTAGTAC 737
Qy 631 IleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArg 650
Db 736 ATAGCGCTGGTGGCAAGGCAATGCTTCAACGAGGAGGAGGATCCCAAAATTAAGCA 677
Qy 651 ThrLeuAspValAspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAsp 670
Db 676 TCTCTTGACGTGGACTAGCTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 617
Qy 671 AspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyLysIleProLysAspIle 690
Db 616 CATATCAACAAGTTCCTGTGGATGGTCCGAATGCGGAGGAGGAGCATCCCAAGGACAT 557
Qy 691 ArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerPro 710
Db 556 AAGAAAGCGATTAATCTTACCGACCGCGGATTAATCAGGTAGATCCGAAGGTGCTCG 497
Qy 711 ThrLeuLeuAsnLysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeu 730
Db 496 GCCCTGTCTCACTCCCTATGTACAAATTAAGCTACTACAGATTCCGGGAATTGAAGTTG 437
Qy 731 AspPheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAsp 750
Db 436 GACTACAGAGCCCATCTGGATATGATCGCACACGTAAAGCCGTCAATGGGATAGGAC 377
Qy 751 IleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIleTyr 770
Db 376 TTGATCTGACCTACCTCGGAGGAGGCTACACACAGAACACTCTGGCTTGTTCGCATCTAT 317
Qy 771 LysValLysAlaProAsp-----AsnArgGluThrLeuAspHisLysProArgValThr 788
Db 316 AGTGTGAAGAGCCGATGATGATCAATAGACCATCTAGAACCAACGAGGAGAGACG--- 260
Qy 789 AsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyr 808
Db 259 -----ATTCTCCAGCAAACTTCATCTCGAAGAAAGAACTCGAAGCGGTCCGAAGGCTAC 206
Qy 809 IleLysAsnLysLeuValPheLysGlyLysLysIleSerLysLysThr 825
Db 205 ATACGAACCCGCGGTGGTGTGTTAAGGGAAACGAACCTTGAATAAACCC 155
```

RESULT 10

US-09-270-767-12856

; Sequence 12856, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12856

; LENGTH: 914

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

US-09-270-767-12856

Alignment Scores:
Pred. No.: 4,79e-72 Length: 914
Score: 742.50 Matches: 140
Percent Similarity: 78.11% Conservative: 17
Best Local Similarity: 69.65% Mismatches: 37
Query Match: 16.87% Indels: 7
DB: 4 Gaps: 2

US-10-028-384-2 (1-826) x US-09-270-767-12856 (1-914)

```
Qy 582 AspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgVal 601
Db 5 GATGACTTCCGCGAGGCTACTACTGCTGCAGATGAACACTCCGAGGAGCGCTCGCATA 64
Qy 602 MetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeuVal 621
Db 65 ATGCTCTGTGGGACTTACGGCTACCAAGATAACGGCCATGCGCAATCGGACGATATTAGT 124
Qy 622 AspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsn 641
Db 125 GATAACAATCTTGGAAACAACAACATATATCGCGCGTGGCCAGGCGATGGCTCTTCG 184
Qy 642 GluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePhe 661
Db 185 GAGGAGAAAAGCCCTACGAGATAATGAGGAACTGGATGTGGACTAGCTTCTCGTGAATTC 244
Qy 662 GlyGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrpMetValArgIle 681
Db 245 GGAGGGCTCACTGGCTACTCATCGACCATATCAACAAGTTCTCTGTGATGGTGGCAT 304
Qy 682 Ala-----GluGlyGluHisProLysAspIleArgGluSerAspTyrPheThr 697
Db 305 GGCGGAGCAGCAGATCGTGGTGGCGAC-----ATCCCGGAAAAAGGACTACTATGCG 355
Qy 698 ProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnLysLeuMet 717
Db 356 GCCAACGAGAGTTCGAGTGGACAAAGGAGGCGCCACCCACACGCTCAATTGTTGATG 415
Qy 718 TyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGly 737
Db 416 TACAAGATGCTACTATCGTTTGGGCAATGTACACGGAAGGTGGCAAGCCCGAGGC 475
Qy 738 PheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGlu 757
Db 476 TAGATCGAGTTTCGTCGCGCGAGATCGCAACAGGACTTTGAATGGATGTCCTGGAG 535
Qy 759 GluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsn 777
Db 536 GAGGCGCTACCAACGAGGACCTGGCTGTGGCTACTACAAAGGTTAAGGATCTCCCGAAT 595
Qy 778 Arg 778
Db 596 CGT 598
```

US-10-028-384-2 (1-826) x US-09-270-767-15440 (1-1660)

QY 553 HisCysThrTrpValThrSerAsnAlaTyrSerSer-----ProSerValValLeu--- 569
Db 713 CATTCG-----CTGCCAACACAGCGCTATGTGACTATTGTCACAGTATTATTCAC 766
QY 570 -----AlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspPheArgGluAla 587
Db 767 TAGCGTCGTTCTGTTCCCAT-----TCCCGCAACATTTAGACGATTTACAGAGGCT 820
QY 588 TyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyr 607
Db 821 TACTACTGGCTTCGCGAAGACACTGCGGATGATGCTCGCGTATGTCITGGTGGGATTAC 880
QY 608 GlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn 627
Db 881 GGATACCAGATAGCGGGAATGGCAACAGACAGACCTAGTGGATAATAATACGTGGAAC 940
QY 628 AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLys 647
Db 941 AATAGTCACATAGCGCTGGTGGTGGCAAGCAATGCTTCAACGAGAGAGAGTCCACGAA 1000
QY 648 IleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyValIleGlyTyr 667
Db 1001 ATTATGACATCTCTTGACGTGGACTAGCTTTTGGTGATCTTTGGGGGTGATCGCTAT 1060
QY 668 SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGluHisPro 687
Db 1061 TCTGGCGATGATATCAACAGATTCCTGTGGATGGTCCGAATTCGCTGAGGAGAGATCCC 1120
QY 688 LysAspIleArgGluSerAspTyrPheThrProGlnGlyLysPheArgValAspLysAla 707
Db 1121 AAGGACATTAAAGAAACGATTAATTACCGACCGGGTGAATTCAGGGGTAGATCCGAA 1180
QY 708 GlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGlu 727
Db 1181 GGTCTCCGGCCCTGCTCAACTGCTTATGTACAAATTAAGTACTTACAGATTCGGGAA 1240
QY 728 MetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGly 747
Db 1241 TTGAAGTTGGACTACAGAGGTCCATCTGGATATGATCGCACAGCTAACCGCTCATGGG 1300
QY 748 AsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuVal 767
Db 1301 AATAAGGACTTCGATCTGACTCTACCTGAGGAGGCGCTACACACAGACACTGGCTTGT 1360
QY 768 ArgIleTyrLysValLysAlaProAsp-----AsnArgGluThrLeuAspHisLysPro 785
Db 1361 CGCATCTATAGGTGAAGAGCGCATGAGTTCAATAGACCATCACTGAAGACCAAGGAG 1420
QY 786 ArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLys 805
Db 1421 AGAAG-----ATTCTCCAGCAAACTTCATTTCAGAAAGAACTCTTAAGGCTCGC 1471
QY 806 ArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLysThr 825
Db 1472 AAGGGCTACATACAGAACCCAGCGGTGTTGTTTAAGGAAACGAACTTGAATAAACCC 1531

RESULT 6

US-09-513-999C-1965
; Sequence 1965, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm

; SEQ ID NO: 1965

; LENGTH: 507

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 122..505

US-09-513-999C-1965

Alignment Scores:

Pred. No.: 1,086-90 Length: 507
Score: 907.00 Matches: 168
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.61% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-513-999C-1965 (1-507)

QY 562 TyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeu 581
Db 2 TACTCTAGTCCCAAGTGTAGTCTCGCCTCATACATCATGATGCGCAGGAATATCTTA 61
QY 582 AspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgVal 601
Db 62 GATGATTATAGAGAGCTTACTTTTGGCTTAGGCAAAATACAGATGAACATGCACAGTA 121
QY 602 MetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuVal 621
Db 122 ATGTCCTGCTGGATATGCTGCTATCAGATAGTGAATGGCTAATAGAACTACGTTGGTG 181
QY 622 AspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsn 641
Db 182 GATATTAACACCTGGATTAACAGCACATAGCAGTGGTGGGAAAGCTATGCTCTTAA 241
QY 642 GluThrAlaAlaTyrIleGlyMetArgThrLeuAspValAspTyrValLeuValIlePhe 661
Db 242 GAAACAGCAGCTATAAAATCATGAGGACTCTAGATGTAGATTATGTTGGTTATTTT 301
QY 662 GlyGlyValIleGlyTyrSerLysAspIleAsnLysPheLeuTrpMetValArgIle 681
Db 302 GGAGGGGTATTGGCTATTCTGCTGATGATATCAAAATTTCTGATGGTTAGGATA 361
QY 682 AlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGlu 701
Db 362 GCTGAAGGAGAAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGGAGAA 421
QY 702 PheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSer 721
Db 422 TTCGGTGTAGACAAAGCAGGATCCCTACTTTTGTGATTCCTTAATAAATGTC 481
QY 722 TyrTyrArgPheGlyGluMetGln 729
Db 482 TACTACAGATTGGAGAAATGCGAG 505

RESULT 7

US-09-248-796A-3089
; Sequence 3089, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208

Qy 202 AlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaCysPheAsp 221
 Db 861 GCGCTTTGATATCCATCGTTCCGGGTATATCTCCGATCCGTCGGGATCGTACGAC 920
 Qy 222 AsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrLeuTrpValIleSerVal 241
 Db 921 AATGAGGCAATCCCATTTTCGATGCTCTTCACCTACTATTTGGGATCAAGCGGTA 980
 Qy 242 LysThrGlySerValPheThrMetCysCysLeuSerTyrPheTyrMetValSer 261
 Db 981 AAGACGGGACGATCTTTGGTGGGTATGTCGGCATTCGGCTACTCTTATATGTCCTCC 1040
 Qy 262 AlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeu 281
 Db 1041 TCGTGGGTGGCTATGCTTCCTGATACCTTAACCTCCCTGCAAGTCTGGCGGTATG 1100
 Qy 282 LeuMetGlnArgTyrSerIleValTyrIleAlaTyrSerThrPheTyrIleValGly 301
 Db 1101 ATCACCGGACGCTTTTCGCAAGGATCTACATAGCATACAGCAGCCTACTACTGCTCGC 1160
 Qy 302 LeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHis 321
 Db 1161 ACCATTCGTGATGAGATCTGTTTGGGATTCACACCCATCCAGAGTCCGACAC 1220
 Qy 322 MetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArg 341
 Db 1221 ATGCTGGCACTGGGAACCTTTGGCCTGTGCCAGATTCAAGCTTTCGTCGACTATCTGGC 1280
 Qy 342 AspArgLeuThrIleGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAla 361
 Db 1281 TCGGCGATCCCAAGGATCATCTTCGATCTGCTCTTCAGACGTTGGTTTCAGTGTGTTG 1340
 Qy 362 GlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGly 375
 Db 1341 ACTGTGTGTTCTGTCGTGGGTACCTCTCTCACGCTTACCGGG 1382

RESULT 4

US-09-270-767-158
 ; Sequence 158, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 158
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-158

Alignment Scores:
 Pred. No.: 2,03e-97 Length: 1660
 Score: 976.00 Matches: 188
 Percent Similarity: 79.29% Conservative: 34
 Best Local Similarity: 67.14% Mismatches: 44
 Query Match: 22.18% Indels: 14
 DB: 4 Gaps: 6

US-10-028-384-2 (1-826) x US-09-270-767-158 (1-1660)

Qy 553 HisCysThrTrpValThrSerAsnAlaTyrSerSer-----ProSerValValLeu--- 569
 Db 713 CATTCG-----CTTGCAACACCGCTATGTGACTATTGTCACGTTATTATCCAC 766
 Qy 570 -----AlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspPheArgGluAla 587
 Db 767 TAGCGTCTCTCTTCTGCTTCCAT-----TCCGGCAACATTTAGACCATTTTCAGAGGCT 820
 Qy 588 TyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyr 607

Db 821 TACTACTGCTTTCCAGAACACCTGCCGATGATGCTCGGTATGTTCTTGGTGGATTAC 880
 Qy 608 GlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn 627
 Db 881 CGATACCATAGTCGGGATGCAACAGACGAGCTAGTGGATAATAATAACGTGGAAC 940
 Qy 628 AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLys 647
 Db 941 AATAGTCACATAGCTGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTACGAA 1000
 Qy 648 IleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyr 667
 Db 1001 ATTATGACATCTCTGACGTGGACTACGTTTGGTGTGATCTTTGGCGGTGTGATCGGTAT 1060
 Qy 668 SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisPro 687
 Db 1061 TCTGGCGATGATCAACAAGTCTCTGTGGATGGTCCGAATGCTGAGGAGAGCATCCC 1120
 Qy 688 LysAspIleArgGlySerAspTyrPheThrProGlnGlyGluPheArgValAspLysAla 707
 Db 1121 AAGGACATTAAAGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGCTAGATGCCAA 1180
 Qy 708 GlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGlu 727
 Db 1181 GGTGCTCCCGCCTCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTTCGGGAA 1240
 Qy 728 MetGlnLeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGly 747
 Db 1241 TTGAAGTTGGACTACAGAGGTCCATCTGGATATGATGCACACGTAACCGCTCATTTGG 1300
 Qy 748 AsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuVal 767
 Db 1301 AATAAGGACTTCGATCTGACCTACCTCGAGGAGGCTACACACAGAACACTGCTTGT 1360
 Qy 768 ArgIleTyrLysValIysAlaProAsp-----AsnArgGluThrLeuAspHisLysPro 785
 Db 1361 CGCATCTATAGGTGAAGAGCGCATGATGATTCATAGACCATCACTGAAGACCAAGAG 1420
 Qy 786 ArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLys 805
 Db 1421 AGAAGC-----ATTCCTCCAGAAACTTCATTTGAGAAAGAACTCTAAGCGTGGC 1471
 Qy 806 ArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLysThr 825
 Db 1472 AAGGCTACATACGAAACCGCGGTTGTTGTTAAGGGAACGAACTTGAATAAACCC 1531

RESULT 5

US-09-270-767-15440
 ; Sequence 15440, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 15440
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-15440

Alignment Scores:
 Pred. No.: 2,03e-97 Length: 1660
 Score: 976.00 Matches: 188
 Percent Similarity: 79.29% Conservative: 34
 Best Local Similarity: 67.14% Mismatches: 44
 Query Match: 22.18% Indels: 14
 DB: 4 Gaps: 6

Thu Dec 16 16:25:06 2004

Db 2110 AAAACCAAGAGGTCCATAAAGAGACCTGAATTA 2142

RESULT 2

US-09-270-767-12331

; Sequence 12331, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12331

; LENGTH: 900

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

US-09-270-767-12331

Alignment Scores:
Pred. No.: 5,09e-118 Length: 900
Score: 1158.00 Matches: 213
Percent Similarity: 90.5% Conservative: 26
Best Local Similarity: 80.6% Mismatches: 25
Query Match: 26.31% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-12331 (1-900)

```
QY 63 SerGlnProAlaGlyTrpGlnSerLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
DB 109 AGCAAGGTGGTGGTACAGAGCTATACCTTCGCCATCTCTAATCGCTGGCTG 168
QY 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
DB 169 GCGGATTTCTCTCGCTTCGCGCTTCGCGCTTCGAGTCGATATCATGAGTTT 228
QY 103 AppProTrpPheAsnTrpArgSerThrHisHisLeuAlaSerHisGlyPheTyrluPhe 122
DB 229 GATCCGTGTTCAACTACCGGCCACCGCTACATCGTGCAGATGGTGTGTACAACTTC 288
QY 123 LeuAsnTrpPheAspGluArgAlaTrpTyProLeuGlyArgIleValGlyGlyThrVal 142
DB 289 CTCACATGTTTCAGCAGCGCGCATGTTATCGCTCGGAGATTTGTGGCGGTACCGTC 348
QY 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsnThrIleAsnIle 162
DB 349 TATCCGGGCTGATGATTACGTCCGCGGGAATCCATTGGCTCTGCACGTACTCAACATA 408
QY 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
DB 409 CCGGTCCATATTCGTGACATCTCGGTTCCTTCCGCGCGCATCTTCAGTGGCTGACCTCC 468
QY 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 202
DB 469 ATCTCCACCTACCTGTCACCAAGAGCTGTGGTCCGCGCGCGCGCTCTTCGCGGCC 528
QY 203 CysPheIleAlaIleValProGlyTyIleSerArgSerValAlaGlySerPheAspAsn 222
DB 529 AGCTTCATCGCATCGTGGCTGCTACATCAGTAGTGGTGGTGGTGGTGGTGGTGGTGG 588
QY 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyTrpValIleSerValIys 242
DB 589 GAGGCGATTCGCATATTCGCTCGAGTTCACCTACTTCTGTGGTGGTGGTGGTGGTGG 648
QY 243 ThrGlySerValPheThrMetCysCysLeuSerTyPheTyTrpMetValSerAla 262
DB 649 ACTGGATCCGTGTTCTGGTCCGCGCGCGCGCTTGTCTTCTTCTTCTTCTTCTTCTT 708
QY 263 TrpGlyGlyTyValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeu 282
DB 709 TGGGTGGCTACGTGTTCATCATCACTGATACCTGATACCTGACGCTCTTCTGTCTCTCT 768
```

```
QY 283 MetGlnArgTyTrpSerLysArgValTyIleAlaTyTrpSerThrPheTyIleValGlyLeu 302
DB 769 ATGGCAGGTATCTCGCGCTCTCTGACAGCTACAGCACTTCTATCATCTGGGACTG 828
QY 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
DB 829 CTGTTCTCATCGCATCCCTTCGTGGGATTCCACCGATACGACCACTGACACATG 888
QY 323 AlaAlaAlaGly 326
DB 889 GCTGCGCTGGGA 900
```

RESULT 3

US-09-270-767-11648

; Sequence 11648, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 11648

; LENGTH: 1386

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

US-09-270-767-11648

Alignment Scores:
Pred. No.: 8,32e-104 Length: 1386
Score: 1032.50 Matches: 196
Percent Similarity: 74.55% Conservative: 53
Best Local Similarity: 58.68% Mismatches: 84
Query Match: 23.46% Indels: 1
DB: 4 Gaps: 1

US-10-028-384-2 (1-826) x US-09-270-767-11648 (1-1386)

```
QY 43 LysAlaAlaGlyGlyAlaAlaProProLysProAlaPro---AlaGlyLeuSerGlyGly 61
DB 381 AAGCCCGCGAGCGGAGAGAGCAAGAAATGACGCTGAGCTGGGGCGGCTGCCAGGGA 440
QY 62 LeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrp 81
DB 441 CTGCTCACCTGGGACAAACAGGAGCACCTGGTCAAGCTGGCCATCTCATCTGCGAGCG 500
QY 82 LeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHisGlu 101
DB 501 GTTTTATCATTTGCCACAGCTTGTCTCTGTGCTCGGATTCGAAAGCGTATCCATGAG 560
QY 102 PheAspProTrpPheAsnTrpArgSerThrHisHisLeuAlaSerHisGlyPheTyrlu 121
DB 561 TTGATCCGTACTTCACTACCCACGACCGCGGTTCTGCGGAGCAGGCGCTTTTACAAG 620
QY 122 PheLeuAsnTrpPheAspGluArgAlaTrpTyProLeuGlyArgIleValGlyGlyThr 141
DB 621 TTCCACAACCTGGTTCGATGACCGCGCTGTGTATCCCTGGCGCGCATCATCGCGGACC 680
QY 142 ValTyProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsnThrLeu 161
DB 681 ATCTATCCCGGCTGATGCTACCTCGGCGCGCTGTACCGCTGATGTGGTGTCTCAAT 740
QY 162 IleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThr 181
DB 741 GTGACCATCGACATACGAAACGCTGTGCTCTTCTGCGCGCTTCTTCTCTCGCTGACC 800
QY 182 SerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 201
DB 801 ACGTGTGTGACCTACGCGCTCACAAAGAGGATACACAGCACTGGAGCTGGACTGGTGGCC 860
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:57:05 ; Search time 160.664 Seconds
(without alignments)
3654.277 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 4401

Sequence: 1 MAFPSAPESKSHKSLNSSPW.....GYKNKLVFKGKISKKTIV 826

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPFO.spool/US10028384/runat_14122004_131520_8052/app_query.fasta_1.3740
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CIGN 1.1 259 @runat 14122004_131520_8052 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2021.5	45.9	2157	4	US-09-614-221A-318
2	1158	26.3	900	4	US-09-270-767-12331
3	1032.5	23.5	1386	4	US-09-270-767-11648
4	976	22.2	1660	4	US-09-270-767-1584
5	976	22.2	1660	4	US-09-270-767-15440
6	907	20.6	507	4	US-09-513-999C-1965
7	831	18.9	867	4	US-09-248-796A-3089
8	806.5	18.3	1660	4	US-09-270-767-1584
9	806.5	18.3	1660	4	US-09-270-767-15440
10	742.5	16.9	914	4	US-09-270-767-12856
11	704	16.0	616	3	US-09-328-111-332
12	595.5	13.5	560	4	US-09-270-767-15600

595.5	13.5	560	4	US-09-270-767-15600	Sequence 16600, A
553	12.6	503	4	US-09-270-767-27262	Sequence 27262, A
544	12.4	594	4	US-09-248-796A-3090	Sequence 3090, Ap
469	10.7	487	3	US-09-385-982-213	Sequence 213, App
427	9.7	245	4	US-09-513-999C-21090	Sequence 21090, A
410	9.3	250	4	US-09-513-999C-1438	Sequence 1438, Ap
406	9.2	487	4	US-09-703-705-1655	Sequence 1655, Ap
406	9.2	487	4	US-09-736-457-1555	Sequence 1555, Ap
406	9.2	487	4	US-09-614-124B-1855	Sequence 1855, Ap
406	9.2	487	4	US-09-671-325-1655	Sequence 1655, Ap
398	9.0	307	4	US-09-702-705-1271	Sequence 1271, Ap
398	9.0	307	4	US-09-702-705-1271	Sequence 1271, Ap
398	9.0	307	4	US-09-736-457-375	Sequence 375, App
398	9.0	307	4	US-09-736-457-1271	Sequence 1271, Ap
398	9.0	307	4	US-09-614-124B-375	Sequence 375, App
398	9.0	307	4	US-09-614-124B-1271	Sequence 1271, Ap
398	9.0	307	4	US-09-671-325-375	Sequence 375, App
398	9.0	307	4	US-09-671-325-1271	Sequence 1271, Ap
398	9.0	307	4	US-09-589-184-375	Sequence 375, App
398	9.0	307	4	US-09-589-184-375	Sequence 375, App
398	9.0	307	4	US-09-658-824-375	Sequence 375, App
398	9.0	307	4	US-09-658-824-1271	Sequence 1271, Ap
392	8.9	307	4	US-09-702-705-588	Sequence 588, App
392	8.9	307	4	US-09-736-457-588	Sequence 588, App
392	8.9	307	4	US-09-614-124B-588	Sequence 588, App
392	8.9	307	4	US-09-671-325-588	Sequence 588, App
392	8.9	307	4	US-09-589-184-588	Sequence 588, App
392	8.9	307	4	US-09-658-824-588	Sequence 588, App
359	8.2	292	4	US-09-313-294A-4442	Sequence 4442, Ap
332	7.5	291	4	US-09-313-294A-4834	Sequence 4834, Ap
326	7.4	414	4	US-09-513-999C-271	Sequence 271, App
297	6.7	302	4	US-09-313-294A-6869	Sequence 6869, Ap
295	6.7	269	4	US-09-313-294A-906	Sequence 906, App

ALIGNMENTS

RESULT 1
US-09-614-221A-318
; Sequence 318, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunananda, Balasulojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 318
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-318

Alignment Scores:			
Pred. No.:	7.21e-213	Length:	2157
Score:	2021.50	Matches:	397
Percent Similarity:	68.5%	Conservative:	118
Best Local Similarity:	52.8%	Mismatches:	183
Query Match:	45.9%	Indels:	53
DB:	4	Gaps:	11

US-10-028-384-2 (1-826) x US-09-614-221A-318 (1-2157)

Qy 68 TTPGINSerLeuLeuserPhetrilleuPhelLeuAlaTrpLeuAlaGlyPheSer 87
Db 34 TTTCAGACCCTTCAGCTGTCATCTCTGTCGCGATTTCGGGCTGCATATCATCA 93

QY 193 TrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIle 212
 DB : : : : :
 DB 439 AAGATTCTAGTGCAGGATTATTGGCAGCTATATTATGGGATTATCCCGAGGTTATATT 498
 QY 213 SerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPhe 232
 DB : : : : :
 DB 499 TCAAGATCAGTGGTGGTCTTATGATTAATGAGCAATGCCATATCTTTATTAATGGCA 558
 QY 233 ThrTyrLeuTrpValLysSerValLysThrGlySerValPheTrpThrMetCysCys 252
 DB : : : : :
 DB 559 ACATTTATTTCTGGATTAAATCAATGAAATGGGTTCACTTTCTATGCCCATTTGACA 618
 QY 253 CysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeu 272
 DB : : : : :
 DB 619 GCATTAATCTATTTCTATATGGTGTAGTGGTGGGATGATGTTTTCATTACCAATTG 678
 QY 273 IleProIleHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArgValTyrIle 292
 DB : : : : :
 DB 679 ATTCATTTACAGTATTGTCTTGATTTTCATGGGTCGTTATAATGCCAACTTTACACT 738
 QY 293 AlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGly 312
 DB : : : : :
 DB 739 GCTTATACATGATGCTGGTGGTACTTGGCATCAATGCAGATTCATTCGTTGGG 798
 QY 313 PheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuGln 332
 DB : : : : :
 DB 799 TTTTACCAATTAGATCAATGATCATATGCTGCATTAGGAGTATTTGGATTGTTACAA 858
 QY 333 AlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeu 352
 DB : : : : :
 DB 859 TTAGTGGCTTTGGTGGTATTGTTGATTAATCAAAAGTTCACCAACCAATTTAAATCATT 918
 QY 353 PhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThr 372
 DB : : : : :
 DB 919 TTGATGATTCTTCACTTGTGTTGGATTAGTATTGGTGGTATTGCTGATTAACA 978
 QY 373 TyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyr 392
 DB : : : : :
 DB 979 GCAATGGTGGTGGTCTCTTGGACAGGTAGATTATTCCTATGGATCAAAATAT 1038
 QY 393 AlaLysIleHisIleProIleAlaSerValSerGluHisGlnProThrTrpVal 412
 DB : : : : :
 DB 1039 GCCAAGATTTCATATTCATATTGCTTCTGTTCTTCAACATCAACCTACTGCTGGCCA 1098
 QY 413 SerPhePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCys 432
 DB : : : : :
 DB 1099 GCATCTTTTTCGATACATAGTATGCTTATTTGGTATTCCCGCTGGTATCTATTATGT 1158
 QY 433 IleLysAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyr 452
 DB : : : : :
 DB 1159 TTCAAGAAATGAAGGATGACACGTTTTCATTTATCATTTTACAGTATTGTTGTTCTAT 1218
 QY 453 PheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSer 472
 DB : : : : :
 DB 1219 TTTCTGCTGTCATGGTAAGATTGATGTGACTTTGACTTCCAGTCAATTGTTGCTGCA 1278
 QY 473 AlaIleAlaPheSerAsnValPheGluHisTyrLeuGly-----Asp 486
 DB : : : : :
 DB 1279 GCAATTCGCTTATCTAAATTTGTTGATGCTATTGTCATTTGTCATTTGTTCACTGAG 1338
 QY 487 AspMetLysArgGluAsnProValGluAspSerSerAspGluAspAspLysArgAsn 506
 DB : : : : :
 DB 1339 AAGTTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398
 QY 507 GlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLys 526
 DB : : : : :
 DB 1399 AGTTCCAGA-----TTTCCAATTTGCTGATATTGCTCA----- 1431
 QY 527 ThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeu 546
 DB : : : : :
 DB 1432 -----AAGTTTGTGTTTACTGACATTTTACATTTTAC 1464

QY 547 LeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSer 566
 DB : : : : :
 DB 1465 CTTTCTTACTTTTGTTTTACATTTGTTGGGTAACATCGAATGCTTATTTTCACTACCATCA 1524
 QY 567 ValValIleAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGlu 586
 DB : : : : :
 DB 1525 GTTGTTTTAGCATCCAGAAACCCAGATGGCTCACACATATCATTTGATTATAGAGAA 1584
 QY 587 AlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTyrTrpAsp 606
 DB : : : : :
 DB 1585 GCTTATTTAGTGGTAAAGTAATGATACACAGAAAGATGCCAAGTTATGCTGTTGGGAT 1644
 QY 607 TyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnThrTrp 626
 DB : : : : :
 DB 1645 TATGGTTATCAATCGGGGTATGGCTGATAGAACCACTTGTGTATAACAATACATGG 1704
 QY 627 AsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyr 646
 DB : : : : :
 DB 1705 AATAACACACATATTGCCACTTGTGTAAGGCAATGCTTCCCTGAAGATGTGCTGAT 1764
 QY 647 LysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGly 666
 DB : : : : :
 DB 1765 GAAATTTTGAGACAAACACGATGTTGATTATGTTTATATATTGGAGGGTTATTGGGT 1824
 QY 667 TyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHis 686
 DB : : : : :
 DB 1825 TATCTGGTGGTATTAACCAATTTCTATGGATGGTAAGAAATGCTGAAGGTTATCTGG 1884
 QY 687 ProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLys 706
 DB : : : : :
 DB 1885 CCTGATGAATCAAGAAAGAGACTACTTACTGACCGAGGAGATATAAAGTGGATAAA 1944
 QY 707 AlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGly 726
 DB : : : : :
 DB 1945 GATGCATCTACTGGCAATGAAGAAATTTCTTGTATGATAAGTTATCTGATCATAGATTCACT 2004
 QY 727 GluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsnAlaGluIle 746
 DB : : : : :
 DB 2005 GAATTTG-----TTTGGAGGTAGAGATGGTGTGATAGAGTTAGAACCAACAATC 2055
 QY 747 ---GlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrp 765
 DB : : : : :
 DB 2056 CCAGCCAAATGAAGTACCGAAATTTGAATGTTGTTGAAGAGCCTTCACATCAGAAATTTGG 2115
 QY 766 LeuValArgIleTyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysPro 785
 DB : : : : :
 DB 2116 ATTGTGAGAAATTACAAAGTTAAAGATTTGGATAATGTTGTTAGAGATTATACATCAA--- 2172
 QY 786 ArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLys 805
 DB : : : : :
 DB 2173 -----GCTACTGCTTTTGAAGAAATCA-----TCATCCGGCACTTCCAAAGAAAC 2217
 QY 806 ArgGlyTyrIleLysAsnLysLeu 813
 DB : : : : :
 DB 2218 AGATCCATAAGAGACCTTAATTTG 2241

Search completed: December 14, 2004, 20:01:33

Job time : 1098.96 secs

Db 1552 ATTGACGATTATCGTGAGGCTTACTACTGCTCGTCAGAAATCTCTCTCAGAACGCCAAA 1611
 Qy 601 ValMetSerTptTptAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
 Db 1612 ATCATGTTCATGCGGATATGCGTATCAATCGTGGCATGCGGACGCCACCTTG 1671
 Qy 621 ValAspAsnThrTptAspAsnSerHisIleAlaLeuValGlyValAlaMetSerSer 640
 Db 1672 GTTGNACACACACTGGACACACACCATATGCTACGGTTGTTAGGCGATGACTCA 1731
 Qy 641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIle 660
 Db 1732 CGCGAGGAAGTCAGTACCCCATCTCCGCCAGCATGATCGATTACGTGCTGGTGG 1791
 Qy 661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
 Db 1792 TTCGGTGTCTGTCAGGTTATCTGGCGATGACATTAACAAATCTTATGGATGGTCCGT 1851
 Qy 681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
 Db 1852 ATCCCGGAAGGTATCTGCGCCGATGAGGTAAAGAGCGGACTCTTTACTGCACCGGT 1911
 Qy 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
 Db 1912 GAATATCTGTCTGAGTGGAGCGACCCCACTATGCGCAACAGCTTGTGTAATAATG 1971
 Qy 721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArg 740
 Db 1972 TCTTATTACAAATTTCACTCTCTC-----TTCCGTCGGGCCCAAGCTGTGCAACCG 2022
 Qy 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisIleGluGluAlaPhe 760
 Db 2023 GTCCGTGGTCAAACTTCCACAGAGGCCCTCAGCTCTCTCATCTCGAAGAGCTTTC 2082
 Qy 761 ThrSerGluHisTptLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
 Db 2083 ACGAGCGAGAACTGATCTGATCTGATCTACAAGGTCAAGGATCTTGACAACTTGGCCGA 2142
 Qy 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
 Db 2143 ---GACCAACACCGGCTGTTCGCT-----TTCCAAAGGCTCTCAAGAAAG 2187
 Qy 801 ThrThrLysArgLysArgGly 807
 Db 2188 CGGAGTACAAAGAGGAAGGA 2208
 RESULT 15
 ID AB232036
 XX AB232036 standard; DNA; 2256 BP.
 AC AB232036;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE Candida albicans essential gene SEQ ID NO 6323.
 KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KW signal transduction; DNA replication; cell division; growth;
 KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
 OS Candida albicans.
 XX
 PN WO200253728-A2.
 PD 11-JUL-2002.
 XX
 PF 26-DEC-2001; 2001WO-US049486.
 PR 29-DEC-2000; 2000US-0259128P.
 PR 20-FEB-2001; 2001US-00792024.
 PR 22-AUG-2001; 2001US-0314050P.
 XX
 PA (ELIT-) ELITRA PHARM INC.

XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
 XX WPI; 2002-566694/60.
 XX P-PSDB; ABP73486.
 XX Constructing strains for identifying gene products as effective targets
 XX for therapeutic intervention, by inactivating in the strain one allele of
 XX a gene and placing other allele of the gene under conditional expression.
 XX Claim 37; SEQ ID NO 6323; 167pp + Sequence Listing; English.
 XX The invention relates to constructing (M1) a strain of diploid fungal
 XX cells in which both alleles of a gene are modified, comprising modifying
 XX one allele by insertion or replacement by a cassette having an
 XX expressible selectable marker and modifying other allele by
 XX recombination, of a promoter replacement fragment with a heterologous
 XX promoter, so that expression of the second allele is regulated by the
 XX promoter. (M1) is useful for constructing a strain of diploid fungal
 XX cells in which both alleles of a gene are modified. The diploid fungal
 XX cells having both alleles modified are useful for identifying a gene that
 XX is essential to the survival or growth of a fungus, a gene that
 XX contributes to the virulence and/or pathogenicity of a fungus, a gene
 XX that contributes to the resistance of a diploid fungus to an antifungal
 XX agent, an antifungal agent that inhibits the growth of a diploid fungus
 XX and for identifying a therapeutic agent for treatment of a mammalian
 XX disease. (M1) is useful for identifying a compound which modulates the
 XX activity of a gene product, preferably enzymatic activity, carbon
 XX compound catabolism, biosynthetic, transporter, transcriptional,
 XX translational, signal transduction, DNA replication and cell division
 XX activity. The method is useful for identifying a compound having the
 XX ability to inhibit growth or proliferation of C. albicans cells and for
 XX treating infection by C. albicans. The present sequence is that of an
 XX essential Candida albicans gene used in the method of the invention.
 XX Note: The sequence data for this patent is not represented in the printed
 XX specification but is based on sequence information supplied to Derwent by
 XX the European Patent Office
 XX SQ Sequence 2256 BP; 648 A; 336 C; 456 G; 816 T; 0 U; 0 Other;

Alignment Scores:
 Pred. NO.: 1,86e-196 Length: 2256
 Score: 2164.00 Matches: 420
 Percent Similarity: 71.39% Conservative: 114
 Best Local Similarity: 56.15% Mismatches: 176
 Query Match: 49.17% Indels: 38
 DB: 6 Gaps: 10
 US-10-028-384-2 (1-826) x AB232036 (1-2256)

Qy 76 IleLeuPheLeuAlaTrpLeuAlaGly-----PheSerSerArgLeuPheAlaValIle 93
 Db 82 ATTATATTATATCG--ATAGCAGGTGCAGCTATTCTCTCGTTTATTTTCGTGATT 138
 Qy 94 ArgPheGluSerIleLeuHisGluPheAspProTrpPheAsnTyrArgSerThrHis 113
 Db 139 CGATTGAAAGTATTATTCATGAATTCGATCTTGGTCAATTCGAGCAACCAATAT 198
 Qy 114 LeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyr 133
 Db 199 TTAGTCACTCATCTCTTTATGAATTTTGAATTTGATGATAGAAC-TGGTACCCA 258
 Qy 134 LeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThrAlaCysLeuIle 153
 Db 259 TTGGGAAGAGTCACTGTTTATATATCCCGGTTTAATGAGTACCTCAGGTGCCATT 318
 Qy 154 ---HisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspValCysValPhe 172
 Db 319 TGGCATATTTTACGTGATTGGTTTGCCTTACCGCTGATATTAGAAATATTGTGTTTA 378
 Qy 173 LeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeu 192
 Db 379 TTAGCACCGATTTCTCGGATTAACGCAATTTGTACTTATTTTGTACTAAAGAAATG 438

QY 266 TyrValPheIleIleAenLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArg 285
 DB 814 TACGTTTTCATCACCACATGATTCATTCGACGCTTTGTTCTCATTCATTCATGCGCAG 873
 QY 286 TyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSer 305
 DB 874 TTCACACACCGCTTATACCGCTTACTCTCTGTAATGTCATTCGAACTATCGCTCC 933
 QY 306 MetGlnIleProPheValGlyPheGlnProIleAargThrSerGluHisMetAlaAla 325
 DB 934 ATGACAGTCCCTTTGTGGATTCCTCCCATCCGAACCTTCGACACATGCGCGCTTG 993
 QY 326 GlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThr 345
 DB 994 GGTGTTTCGTCCTTGTACAGTCATCGGATTCGTGCAAGTCGTCGACATCGTGCT 1053
 QY 346 LysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPhe 365
 DB 1054 GCGAAGCAATTCAGCTCTCTCAAGCTTTTGTGCGCGCTTATTCGCTCGATTTT 1113
 QY 366 LeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProIlePheArgPheTyr 385
 DB 1114 CTGCTCCCTGTCACCTTGTGCTTCTGATGATCGCCCTTCGCTGGAAGATTTAT 1173
 QY 386 SerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGlu 405
 DB 1174 TCTCTTTGGATCTGCGTATGCGAGGTCCACATGCTTATTCCTCGCTCCGAA 1233
 QY 406 HisGlnProThrThrTyrPheValSerPhePheAspLeuHisIleLeuValCysThrPhe 425
 DB 1234 CACGAGCCACCGCTTGGCTCATCTTACTTGACCTCGAATGCTTATCTCTTTTC 1293
 QY 426 ProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAlaLeu 445
 DB 1294 CTGCGCGTGTCTCTGCTGTTTCAAGAGCTTTCGCGATGAGAGATCTTCATCATCT 1353
 QY 446 TyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThr 465
 DB 1354 TATCGCGTTCATGAGCTTATTTGCGGCTGTCATGCTGCTGCTGCTGCTGCTGCTGCT 1413
 QY 466 ProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGly 485
 DB 1414 CTTGTTGTCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1470
 QY 486 AspAspMetLysArgGluAsnProProValGluAspSerSerAspGluAspAspLysArg 505
 DB 1471 -----GACCCGCTCATCCCCGAAAGCGACGAGGAGCTGCGAG 1509
 QY 506 AsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGlu 525
 DB 1510 TCTCAGCGCAGGTGTCTCCAAAGTCC---AAGCGGAAGAAGATGCGCGCTGCCACGCC 1566
 QY 526 -----LysThrGluGluGlyLeu--- 531
 DB 1567 AATAAGACGGGTTCTCTTTCACAGGTATTTTGGCGGCAAGTCTGCTCGGCATCTTT 1626
 QY 532 GlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAla 551
 DB 1627 GGTCTCGACATCGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1686
 QY 552 ValHisCysThrThrTyrValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSer 571
 DB 1687 CTTCACTGCATATGTGACTTCAACAGCGTATTTCTTCGCTTCAGTGTGCTGCTGCTGCTGCTGCTGCT 1746
 QY 572 TyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeu 591
 DB 1747 CGAAACCGGATGTAGCCAAATATCATTTGATGATTTCCGAGAGGCTTACTACTGATTT 1806
 QY 592 ArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIle 611
 DB 1807 CGCCAAACACCGCGAGACGAGCGTCATGCTCTGTTGGGATTCAGCGTACCGATC 1866
 QY 612 AlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 631

DB 1867 GCTGGTATGCTGATCGCCACCCCTTGTGTATAACAATCTGTAATAACCCACAT 1926
 QY 632 AlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr 651
 DB 1927 GCCACAGTTGGTAAGCCATGCTTCCACGAAGATGCGCATATCTTCTGAGGAAG 1986
 QY 652 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 671
 DB 1987 CATGATGTGATACCTTCTTGTGATCTTGGGGCTTATTTGGGCTACTCTGGTGACAT 2046
 QY 672 IleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArg 691
 DB 2047 ATCAACAAGTTTGTGGATGTTAGGATCTCACAAGGTGAATGCGCTGACGAGTGCGAG 2106
 QY 692 GluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThr 711
 DB 2107 GAAGTCAACTACTTACTCAAGAGGGAGTATGCTGTCGATGACGAGGCCACCCCTACT 2166
 QY 712 LeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAsp 731
 DB 2167 ATGAAGAATCTCTCATGTACAAATGTCTTACTACCGCTTCCCGAGCTT----- 2217
 QY 732 PheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIle 751
 DB 2218 TATGGTGGACACCGGCTCAAGACAGGTTCCAGGCCAAATATATCCCTCAACAGTGT 2277
 QY 752 LysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLys 771
 DB 2278 ACTCTTGATACTCTTACGAGAGGTTTACATCCGAAATTTGGATCGTCAGGATCTTACAAG 2337
 QY 772 ValLysAlaProAsp-----AsnArgGluThrLeuAspHisLys 784
 DB 2338 GTCAAGAAGAAGATCCCATTCGACGA-----GACCAACAAG 2373
 RESULT 14
 ABT20789
 ID ABT20789 standard; DNA; 2232 BP.
 AC ABT20789;
 XX
 DT 16-APR-2003 (first entry)
 XX
 DE Aspergillus fumigatus essential gene #3147.
 XX
 KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 XX cancer; contamination; biofilm; antibody; immune response; ds.
 OS Aspergillus fumigatus.
 XX
 PN WO200286090-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 23-APR-2002; 2002WO-US013142.
 XX
 PR 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 DR
 PT New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX
 PS Disclosure; Page; 175pp; English.

XX 20-FEB-2001; 2001WO-US005171.
 XX 17-FEB-2000; 2000US-0183319P.
 XX 16-MAR-2000; 2000US-0189862P.
 XX 25-MAY-2000; 2000US-0207454P.
 XX 09-JUN-2000; 2000US-0211314P.
 XX 18-JUL-2000; 2000US-0219007P.
 XX 13-DEC-2000; 2000US-0255281P.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX Schlegel R, Endege WO, Monahan JE;
 XX WPI; 2001-662795/76.
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 XX prostate cells and correlating with presence of prostate cancer, useful
 XX for detecting presence of prostate cancer, stage of prostate cancer.
 XX Claim 1; Page 4632-4633; 11750pp; English.
 XX The invention relates to an isolated nucleic acid molecule (I) comprising
 XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 XX specification or its complement. (I) is useful for: (a) assessing whether
 XX a patient is afflicted with prostate cancer; (b) monitoring the
 XX progression of prostate cancer in a patient; (c) assessing the efficacy
 XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
 XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
 XX (e) selecting a composition for inhibiting prostate cancer in a patient;
 XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 XX determining whether prostate cancer has metastasized in a patient; (h)
 XX assessing the aggressiveness or indolence of prostate cancer in a patient
 XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 XX Sequence 2760 BP; 688 A; 645 C; 658 G; 769 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 3,24e-203 Length: 2760
 Score: 2236.00 Matches: 429
 Percent Similarity: 74.20% Conservative: 103
 Best Local Similarity: 59.83% Mismatches: 147
 Query Match: 50.81% Indels: 38
 DB: 5 Gaps: 7

US-10-028-384-2 (1-826) x ABV24502 (1-2760)

QY 70 SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerArgLeu 89
 DB 150 ACACUUTTGAGCTTCTCATTTCTGTCATGGCTGCTGTTATTCCTTCCTCCACTGTCG 209
 QY 90 -PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrp 109
 DB 210 TTTTGTGCTGCTGAGATTGAAGTTGAAGTTGAAGTTGAAGTTGAAGTTGAAGTTGAAGTT 269
 QY 109 gSerThrHisHisLeuAlaSerHisGlyPheTyrrGluPheLeuAsnTrpPheAspGlu 129
 DB 270 GACTACCAAGTTTCCTGGCTGAGGAGGGGTTTATTAATTCATTAATTCATTAATTCAT 329
 QY 129 gAlaTrpTrpProLeuGlyArgIleValGlyGlyThrValTyrrProGlyLeuMetIle 149
 DB 330 AGCTTGGTACCTTTGGACGAGATCATTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 389
 QY 149 rAlaGlyLeuIleHisTrpIleLeuAsn-ThrLeuAsnIleThrValHisIleArgAsp 169
 DB 390 CTCGTGCTGAATCTACCATGACTCCATTTTTCACATCACCATTCGCAATTCGGAATG 449
 QY 169 aICysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeu 189
 DB 450 TCTGTGTGTCTGCGCCCT 509
 QY 189 hArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaValP 209

Db 510 CCAAGAGCTCAAGGATCGAGGGCTGGCTTCTTCTGCTGCCATGATTCGTAGTTC 569
 QY 209 roGlyTyrrIleSerArgSerValIleAlaGlySerPheAsnGluGlyIleAlaIlePheA 229
 Db 570 CTGGATATATCTCCGACTCTGGCTGCTCTATGATTAATGAAGGATTCGCATCTTTT 639
 QY 229 lAlaLeuGlnPheThrTrpTyrrLeuTrpValIleSerValIleSerValIleSerVal 249
 Db 630 GCATGCTACTACCTACTACATGGATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 689
 QY 249 hrMetCysCysCysLeuSerTyrrPheTyrrMetValSerAlaTrpGlyGlyValPheI 269
 Db 690 CAGCTAAGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 749
 QY 269 lAlaLeuLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrrSerIys 289
 Db 750 TGAATCACTTAATTCCTCTCCAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 809
 QY 289 rGValTyrrIleAlaTyrrSerThrPheTyrrIleValGlyLeuIleLeuSerMetGlnIle 309
 Db 810 GGATCTATGTGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 869
 QY 309 roPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheA 329
 Db 870 CTTTGTGGGTTTCCAGCTGCTCTTTCATCAGAGCATGCGCAGCTTTGGGGTCTTTG 929
 QY 329 lAlaLeuGlnAlaTyrrAlaPheLeuGlnTyrrLeuArgAspArgLeuThrLysGlnGlu 349
 Db 930 GTCTTGCAGATCCTGATGCTTGTGGATTACCTGCGCAGCAAGTTGAATCCACCAAT 989
 QY 349 hGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerVal 368
 Db 990 TTGAAGTTCTTTCCGAGGAGGCTCTCTCTG---GTAGGCTTTTGTCTCTCTCTCTC 1046
 QY 369 --IleTyrrLeuThrTrpTyrrGlyTyrrIleAlaProTrpSerGlyArgPheTyrrSer 388
 Db 1047 GAGCTCTCTCATGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1106
 QY 388 rpAspThrGlyTyrrAlaIleHisIleProIleAlaSerValSerGluHisGlnP 408
 Db 1107 TGGATCCCTCTTATGCTAAGAACCAACATCCCATCCTTCTGCTGCTGAGCATCAGC 1166
 QY 408 roThrTrpValSerPhePhePheAspLeuHisIleLeuValCysThrPheProAlaG 428
 Db 1167 CCACAACCTGCTCTCATACTATTTGACCTGCAGCTCTCTCTCTCTCTCTCTCTCT 1226
 QY 428 lyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAlaLeuTyrrAla 448
 Db 1227 GCCTCTTACTGCTTTAGCAACCTGCTGATGCCGAGATTTTATCATCATGATGCTG 1286
 QY 448 leSerAlaValTyrrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProVal 468
 Db 1287 TGACCAAGATGATCTTTTCAGCTGTAATGGTGGCTTAATGCTAGTGTGGCACCTGTA 1346
 QY 468 alCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrrLeuGlyAsp--A 487
 Db 1347 TGTGCTATCTCTGTCATTTGAGTCTCCAGGCTCTCTCCACATACATGAAGATCTGG 1406
 QY 487 spMetLysArgGluAsnProProValGluAspSerSerAspGluAspLysArgAsn 507
 Db 1407 ACATAAGTCTGCTCA----- 1420
 QY 507 lGlyAsnLeuTyrrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLys 527
 Db 1421 -----GACAAAGAGAGCAAGCAAGGATTCACCTACCTACCTATTATGA 1466
 QY 527 hrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuLeu 547
 Db 1467 ATGAA-----GTGCAAGTGGGATGATGATGATGATGATGATGATGATGATGATG 1508
 QY 547 enMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrrSerSerProSer 567
 Db 1509 TCATACCTACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1568

[illegible]

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2004-204755/20.
 DR P-PSDB; ADL30662.
 XX
 PT New oligonucleotide primers (830 cDNAs) useful for synthesizing full
 PT length human cDNAs.
 XX
 PS Example 1; SEQ ID NO 2694; 1340bp; English.
 XX
 CC This invention relates to a novel primers useful for synthesizing full
 CC length cDNA molecules that encode human proteins. Specifically, it refers
 CC to secretory or membrane proteins that are potential therapeutic agents/
 CC target molecules in the field of medicine, and in particular genes
 CC encoding proteins that are associated with signal transduction,
 CC glycoproteins and transcription. The present invention describes a method
 CC for efficiently cloning a full length human cDNA from both the 5' and 3'
 CC ends using the oligo-capping method. This polynucleotide sequence is a
 CC full length human cDNA clone of the invention.
 XX
 SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,26-216 Length: 1664
 Score: 2370.00 Matches: 443
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 53.85% Indels: 0
 DB: 12 Gaps: 0

US-10-028-384-2 (1-826) x ADL30661 (1-1664)

Qy 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTrp 20
 Db 28 ATGGCGGAGCCCTCGGCGCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGTGG 87
 Qy 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 Db 88 AGTGCCCTCATGGCCCTGGGAAACAGCCGCGAGCCAGCCAGCCGCGGCGGCGGCGGCGG 147
 Qy 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 Db 148 GGCACAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 207
 Qy 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuSerSerPheThrLeuPheLeuAla 80
 Db 208 GGGCTGTCCAGCGCGCTGGGTGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 267
 Qy 91 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValAlaArgPheGluSerLeuHis 100
 Db 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGTGCATCGCTTCGAAAGCATCATCCAC 327
 Qy 101 GluPheAspProTrpPheAsnTyArgSerThrHisLysLeuAlaSerHisGlyPheTy 120
 Db 328 GAGTTCGACCGCGTGGTAACTAGATCAACACATCATCTTCATCTCATGGGTTCAT 387
 Qy 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgileValGlyGly 140
 Db 388 GAATTTTAAATTTGGTTTATGATGAAGAGCATGGTATCCATAGGAAGAATAGTAGGTGT 447
 Qy 141 ThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
 Db 448 ACTGTTTACCAGGCTTATGATCAACCGCTGGCGCTTATTCATGGATTTTAAATACATG 507
 Qy 161 AsnIleThrValHisLysAlaAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 Db 508 AACATACTGTTCATATAAGAGAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 567
 Qy 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
 Db 568 ACATCTATATCT 627

Qy 201 AlaAlaCysPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySerPhe 220
 Db 628 GCTGCTGT 687
 Qy 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpValValSer 240
 Db 688 GATAAAGAGGCAATGCTATTTTGCAGTTCAGTTCACATCTATTATGGGTAAATCT 747
 Qy 241 ValIleThrGlySerValPheTrpThrMetCysCysLeuSerTyPheTyMetVal 260
 Db 748 GTAAAGACCTGGGTGAGTTTTTGGCAATGTGCTGCTCTATCTATCTATCTATCTAT 807
 Qy 261 SerAlaTrpGlyGlyTrpValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
 Db 808 TCTGCTTGGGTGGTGTATGATTATCATCAATCTTATCCAGTGCATGATTGTGTG 867
 Qy 281 LeuLeuMetGluArgTySerLysArgValTrpIleAlaTySerThrPheTyIleVal 300
 Db 868 TTACTGATGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATGTG 927
 Qy 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 Db 928 GGTTTAATATTATCAATGCAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTGAA 987
 Qy 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyAlaPheLeuGlnTyLeu 340
 Db 988 CACATGCGAGCTGCAGGTGTCTTTCATTTGCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1047
 Qy 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 Db 1048 AGAGACCGATTAAACAAACAGAGTTCAGAGCCCTTTCTTTTGGGTGATCACTAGCT 1107
 Qy 361 AlaGlyAlaValPheLeuSerValIleTyLeuThrTrpThrGlyTrpIleAlaProTrp 380
 Db 1108 GCAGGTGCTGTGTCTTCTAGTCTCATCTATTATGCTTATACAGGTACATTCACCATG 1167
 Qy 381 SerGlyArgPheTySerLeuTrpAspThrGlyTrpAlaLysIleHisIleProIleIle 400
 Db 1168 AGTGGCGAGCTTTTATTCATTTGGGTATCGGTGGTGGTGGTGGTGGTGGTGGTGG 1227
 Qy 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePhePheAspLeuHisIle 420
 Db 1228 GCATAGTGTCTGAGCATCAACCTAGCTTGGGTGTCTTCTTCTTGTATCTACATATT 1287
 Qy 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
 Db 1288 CTTGTATGTACCTTCCAGCAGGCTTTGGTTCGCATCAAAAATATCAACGATGAAGA 1347
 Qy 441 ValPheValAla 444
 Db 1348 GTATTTGGTGTCT 1359
 RESULT 8
 ABL02795
 ID ABL02795 standard; cDNA; 2855 BP.
 AC ABL02795;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2867.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 XX
 OS pharmaceutical; gene; ss.
 XX
 FN Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US009231.
 XX

XX The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a full length human cDNA of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in CD-ROM format directly
 CC from EPO

XX SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;

Alignment Scores: Pred. No.: 2,2e-216 Length: 1664
 Score: 2370.00 Matches: 443
 Percent Similarity: 93.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 53.83% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x AAK94164 (1-1664)

QY	1	MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuSerProTyr	20
DB	28	ATGGCGGAGCCCTCGCCCGGAGNAGCAAGTCGTCCTCAACTCGTCCCGTGG	87
QY	21	SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys	40
DB	88	AGTGGCGCTCATGGCCCTGGGAAACAGCGCGGACGCGCCGCGGCGCGGCGCCAGTGC	147
QY	41	AlaHisLysAlaAlaGlyValAlaAlaProProLysProAlaProAlaGlyLeuSerGly	60
DB	148	GGCGACAGCG	207
QY	61	GlyLeuSerGlnProAlaGlyTyrGlnSerLeuLeuSerPheThrLeuPheLeuAla	80
DB	208	GGCGCTGTCCAGCGCGCGTGGTGGCGAGTCGCTCTCTCTCTCTCTCTCTCTCTCTCT	267
QY	81	TyrLeuAlaGlyPheSerSerArgLeuPheAlaValLeuArgPheGluSerIleHis	100
DB	268	TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTCGGAGCATCATCAC	327
QY	101	GluPheAspProTyrPheAsnTyrArgSerThrHisLysLeuAlaSerHisGlyPheTyr	120
DB	328	GAGTTCGACCCCTGGTAACTATATAGATCAACATCATCTTCATCTCATCTCGGTTCT	387
QY	121	GluPheLeuAsnTyrPheAspGluArgAlaTyrTyrProLeuGlyArgIleValGlyGly	140
DB	388	GAATTTTAAATTTGGTTTGGTGAAGAGCATGGTATCCACTAGGAGAAATAGTAGTGT	447
QY	141	ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisThrIleLeuAsnThrLeu	160
DB	448	ACTGTTTACCAGGTTGATGATAACCGCTGGCTTATTCATTGGATTTTAAATATACATG	507
QY	161	AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu	180
DB	508	AACATAAATGTTCACATAGAGAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	567
QY	181	ThrSerIleSerThrPheLeuLeuThrArgGlnLeuThrPheGlnGlyAlaGlyLeuLeu	200
DB	568	ACATCTATATCTACTTCT	627
QY	201	AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe	220
DB	628	GCTGCTGTGTTTATGCTATTCTACAGGCTACATCTCGTTCAGTAGTGTGATCCTTT	687
QY	221	AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValLysSer	240
DB	688	GATATGAAGGCATTCGTATTTTTCGACTTCAGTTTCATACATCTATTATTCGGTAAATCT	747

QY	241	ValIleThrGlySerValPheThrPheThrMetCysCysLeuSerTyrPheTyrMetVal	260
DB	748	GTAAACACCTGGGTCAGTTTTTTGGACAATGTGCTGCTGCTTATCTATCTATCTATCT	807
QY	261	SerAlaTyrPheGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu	280
DB	808	TCGCTTGGGCTGTTATGTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT	867
QY	281	LeuLeuMetClnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal	300
DB	868	TTTACTGATGAGAGATACAGCAAGAGTCTACATAGCATATAGCATCTTCTACATTTTG	927
QY	301	GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu	320
DB	928	GGTTTAAATATTAATCAATGCAGATACCTTTTGTGGGATTCAGCCCAATCAGAACAGTGA	987
QY	321	HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu	340
DB	988	CACATGGCAGTGCAGGTGCTTTGCAATGTGCTGCAAGTTATGCTTCTTTCAGTATCTG	1047
QY	341	ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla	360
DB	1048	AGAGACCGATTAAACAAACAGAGTCCAGACCTTTTCTTTTGGGTGTATCATCTAGCT	1107
QY	361	AlaGlyValAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyr	380
DB	1108	GCAGTGTGCTGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1167
QY	381	SerGlyArgPheTyrSerLeuTyrPheThrGlyTyrAlaLysIleHisIleProIleIle	400
DB	1168	AGTGGCAGGTTTTTATTCATTGTGGGATCTGGGTATGCAGGTTATACAGTTTCAATATT	1227
QY	401	AlaSerValSerGluHisGlnProThrThrTyrValSerPhePheAspLeuHisIle	420
DB	1228	GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGTCTTCTTCTTCTTCTTCTTCTTCT	1287
QY	421	LeuValCysThrPheProAlaGlyLeuTyrPheCysIleLysAsnIleAsnAspGluArg	440
DB	1288	CTTGTATGTACCTTCCAGCAGCGCTTTGGTTCTGCTCAAAATATCAACGATGAAGA	1347
QY	441	ValPheValAla 444	
DB	1348	GTATTGGTGTCT 1359	
RESULT 7			
ID	ADL30661	standard; cDNA; 1664 BP.	
XX	ADL30661		
AC	ADL30661		
DT	20-MAY-2004	(first entry)	
XX	20-MAY-2004		
DE	Full length human cDNA clone SeqID 2694.		
XX	human; medicine; signal transduction; glycoprotein; transcription;		
KW	oligo-capping method; ss; gene.		
XX	Homo sapiens.		
OS	EP1396543-A2.		
XX	10-MAR-2004.		
PN	07-JUL-2000; 2003EP-00025638.		
PD	08-JUL-1999; 99JP-00194486.		
XX	11-JAN-2000; 2000JP-00118774.		
PR	02-MAY-2000; 2000JP-00183865.		
XX	07-JUL-2000; 2000EP-00114089.		
XX	(REAS-) RES ASSOC BIOTECHNOLOGY.		
PA			
XX			

Db 2916 GCGCTCTTTGGCCGCTGGTGGTCTCACCATGCTGGCGGTTGTGGCCCGTGGAGTGA 2857
 Qy 383 ArgPheTyrSerLeuThrAspThrClyTyrAlaLysIleHisIlePheIleLeuAsp 402
 Db 2856 CCGTCTCTACTCGCTGGGATCTGGCTACGCCAAGATCCACATTCCTCATCTGGATCC 2797
 Qy 403 ValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIleLeuVal 422
 Db 2796 GTGTGGAGCATCAGCCACCACTGGTCTCGTCTCTCTTCTGATCTGCACATCTCTGGT 2737
 Qy 423 CysThrPheProAlaGlyLeuThrPheCysIleLysIleAsnIleAsnAspGluArgValPhe 442
 Db 2736 TGGCGCTTCCAGTGGAGTGGTGGTACGATCAGCAGATCAACGACGAGCGGTTC 2677
 Qy 443 ValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 462
 Db 2676 GTGGTGTCTGACGCCATCAGTCCGCTTACTTCGCTGCTGATGGTGGCTTGTGATGTG 2617
 Qy 463 ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482
 Db 2616 ACCCTCAGCCCGTGGTGTGATGCTGGCCGAGTGGCCCTTTTCGGAGCTGTGGATGTG 2557
 Qy 483 TyrLeuGlyAspAspMetLysArg-----GluAsn 492
 Db 2556 TTCCTGCAAGAGATTCGCTTAAGCGAATGGGCACAGCCATAAGCGCAGCCACCGAAGTG 2497
 Qy 493 ProProValGluAspSerSerAspGluAspLysArgAsnGlnGlyAsnLeu----- 510
 Db 2496 GATGAAGCTGAGGATTCATTTGA-GAAGAGACGCTGTACGACAAGTGAGTTCCTACTA 2438
 Qy 511 -----TyrAspLysIleGlyLysValArg 518
 Db 2437 ACACATCCATGGGATTTGATTTTAAATATACATCGCATTTGCAGCGTGGCAAGCTG--- 2381
 Qy 519 LysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsnIleLysSer 537
 Db 2380 AAGCATCGTACTAAGCATGATGCCAGCAGGATACGGCTGACGCTCCACCTGAGAGT 2321
 Qy 538 IleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
 Db 2320 ATTGTATTATTGGCGGTTCATATCTGTGTGATGATGTTCGCTGCCACTGCCAGTGGTG 2261
 Qy 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
 Db 2260 ACCAGCAATGCTACTCCAGTCCCTCCATGCTGTCTGTGCTTCCACACAGTCAAGATGG 2201
 Qy 577 -----Thr 577
 Db 2200 TAAGCACCAAGTAGCATCATGCTCTTTAAACCTATCTAACCTTTTCTCCACTAGATCC 2141
 Qy 578 ArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGlu 597
 Db 2140 CGCAACATTTAGACGATTTACAGAGAGCTTACTACTGCTTTCCGACAGACATCCCGAT 2081
 Qy 598 HisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 617
 Db 2080 GATGCTCGCGTATGCTCTGTGGGATACGGATACAGATACGGGAATGGCAACAGA 2021
 Qy 618 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 637
 Db 2020 ACGACGCTAGTGGATAATAATACGTGGAACAATAGTCACATAGCTGCTGTGGTGGCAAGCA 1961
 Qy 638 MetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTrpVal 657
 Db 1960 ATGCTTTCAACCGAGGAGAGTCTCTAGAAATATGACATCTCTGACGTGGACATCGT 1901
 Qy 658 LeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 677
 Db 1900 TTGGTGTACTTTGGCGGTGATCGGCTATTCCTGGCATGATATCAACAAGTCTCTGTGG 1841
 Qy 678 MetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThr 697
 Db 1840 ATGGTCCGANTGTGAGGGAGAGCATCCCAAGGACATTAAGGAAGCGATTACTTTACC 1781

Qy 698 ProGlnGlyLeuPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMet 717
 Db 1780 GACCGGGTGAAATCAGGTAGATGCCAAGGTGCTCCGCCCTGCTCACTGCTTATG 1721
 Qy 718 TyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThr----- 734
 Db 1720 TACAAATTAAGCTACTACAGATTCGGGAATTGAAGTTGGACTACAGTA-AGCGGAAAC 1662
 Qy 735 -----ProProGlyPheAsp 739
 Db 1661 ATTTCTCTCAGGTAGGATGCACACTAACTGCCTGTTTCAGAGTCCATCTGGATATGAT 1602
 Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAla 759
 Db 1601 CGCACACGTAAAGCCGCTCATTTGGGAATAAGGACTTCGATCTGACCTACCTGGAGAGGCC 1542
 Qy 760 PheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAsp-----Asn 777
 Db 1541 TACACACAGAACACTGGCTTGTTCGATCTATAGGTGAAGAACCGCATGAGTTCAAT 1482
 Qy 778 ArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeu 797
 Db 1481 AGACCATCTACTGAAGACCAAGGAGAGAACG-----ATTCTCCAGCAAACTTCATT 1431
 Qy 798 SerLysLys----- 800
 Db 1430 TCAGAAAGTAGTGTACCTCCAGCTCCCAATGNACTTCAATTAACATCTGATCTTTG 1371
 Qy 801 ---ThrTyrArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLys 819
 Db 1370 CAGAATCTTAAGCGTCGAAAGGCTTACATACGAAACCGCGTGTGTGTTAAGGGGAAA 1311
 Qy 820 LysIleSerLysLysThr 825
 Db 1310 CGAACCTTGAAATAAAC 1293
 RESULT 6
 AAK94164
 ID AAK94164 standard; cDNA; 1664 BP.
 XX AAK94164;
 AC AAK94164;
 DT 06-NOV-2001 (first entry)
 XX Human full-length cDNA, SEQ ID NO: 2694.
 DE Human full-length cDNA; cDNA synthesis; oligo-capping; ss.
 XX Homo sapiens.
 OS Homo sapiens.
 XX EPI130094-A2.
 XX 05-SEP-2001.
 XX 07-JUL-2000; 2000EP-00114089.
 XX 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI; 2001-524255/58.
 DR P-PSDB; AAM93252.
 XX 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 PS Claim 8; SEQ ID NO 2694; 1380pp + Sequence Listing; English.

QY 717 MetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProPro 736
 Db 2084 ARGTAACAATTAAGCTACTACAGATTCCGGGAATTTGAAGTTGGACTACAGAGTCCATCT 2143
 QY 737 GlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeu 756
 Db 2144 GGTATGATCGCACACGTAACCGCTCATTTGGGAATAAGGACTTCGATCTACCTACCTG 2203
 QY 757 GluGluAlaPheThrSerGluHisTyrPheValArgIleTyrLysValLysAlaProAsp 776
 Db 2204 GAGGAGCGCTACACACAGAACTCTGGCTTGTTCGCATCTATAGGTGAAGAGCGCAT 2263
 QY 777 -----AsnArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGln 794
 Db 2264 GAGTTCAATAGACCATCACTCAAGACCAAGGAGAGAAGC-----ATTCTCCAGCA 2314
 QY 795 LysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuVal 814
 Db 2315 AACITCATTTGAGAAAGAACTCTAAGCGTCGCAAGGCTCATATACAAACCGCGTT 2374
 QY 815 PheLysLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 825
 Db 2375 GTTGTTAAGGAAACCAACCACTTGAATAAACC 2407

RESULT 5

ABL13246/c
 ID ABL13246 standard; cDNA; 4922 BP.

AC ABL13246;
 XX

DT 26-MAR-2002 (first entry)
 XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 XX

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW Pharmaceutical; gene; ss.

OS Drosophila melanogaster.
 XX

PN WO200171042-A2.
 XX

PD 27-SEP-2001.
 XX

PF 23-MAR-2001; 2001WO-US009231.
 XX

PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.
 XX

PI Venter JC, Adams M, Li PWD, Myers EW;
 XX

XX WPI; 2001-65860/75.
 DR P-PSDB; ABB69143.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.

XX Claim 1; SEQ ID NO 34220; 21bp + Sequence Listing; English.
 PS

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175), and the encoded proteins (ABBS7737-
 CC ABBS2072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Alignment Scores:
 Pred. No.: 1.8e-260 Length: 4922
 Score: 2838.00 Matches: 557
 Percent Similarity: 76.24% Conservative: 88
 Best Local Similarity: 65.84% Mismatches: 114
 Query Match: 64.49% Indels: 89
 DB: 4 Gaps: 10

US-10-028-384-2 (1-826) x ABL13246 (1-4922)

QY 63 SerGlnProAlaGlyTyrGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
 Db 3816 AGCAAGGTGGTGGCTACAGACGCTTAATCACTTCGCCATCTCTGTAATGCTGGCTG 3757
 QY 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
 Db 3756 GCGGATTTTCCTCTCCCTCTTCGCGGTCATCCGTTTCGATCGCATTAATCATGAGTT 3697
 QY 103 AspProTyrPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyrGluPhe 122
 Db 3696 GATCCGTGGTTCAACTACCGGGCCACCGCTACATGGTGCAGATGGTGGTACAACTTC 3637
 QY 123 LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyGlyThrVal 142
 Db 3636 CTCAACTGGTTCGACGAGCGCATGGTAATCCCTCGGAGGATTTGGGCGGTACCGTCC 3577
 QY 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsnThrLeuAsnIle 162
 Db 3576 TATCCCGGCTGATGATTACGTCGCGGGAATCCATTCGCTGCTGCAGTACTCAACATA 3517
 QY 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 Db 3516 CCGGTTCATATTCGTGACATCTCGGTGTTCTTGGCGCGCATCTTCAGTGGCTGACCTCC 3457
 QY 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAlaGlyLeuAlaAla 202
 Db 3456 ATCTCCACCTACCTGCTGACCAAGGAGCTGTGGTCCGCGGCGCGGCTCTTCGCCGCC 3397
 QY 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn 222
 Db 3396 AGTTTCATCGCATCTGCTGCTGCTACATCATGATGATGGTGGTGGTGGTGGTGGTGG 3337
 QY 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLys 242
 Db 3336 GAGGCGCATTCGCATATTCGCCCTGCAGTTCACCTACTTCTTCTGGGTGGTGGTGGTGG 3277
 QY 243 ThrGlySerValPheThrThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
 Db 3276 ACTGGATCCGTGTCTGGTGGCGCGCGGCTTTGTCTTCTTCTATCATGGTGTCCGCC 3217
 QY 263 TrpGlyGlyTyrValPheIleIleLeuLeuIleProLeuHisValPheValLeuLeuLeu 282
 Db 3216 TGGGGTGGGTACGTTTCATCATCAACCTGTATACCCCTGCAGCTCTTCTGCTACTGCTCAT 3157
 QY 283 MetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
 Db 3156 ATGGGAGGTACTTCGCGCGGTCTGGTGCACCATACGACCTTCTATCTCTGAGACTG 3097
 QY 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 Db 3096 CTGTTCTCCATGCAGATCCCTTCGTGGGATTCACACCGGATACGACCGAGTGAACACATG 3037
 QY 323 AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
 Db 3036 GCTTGGCGGTGGAGTGTGTGTCTTATGGCGCGTGGCCACTTGGCGCATTTGAGTCC 2977
 QY 343 ArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaGly 362
 Db 2976 GTGCTTCGCGCAACGAGTTCGGAAGTGTTCATCTGTCGCGGAGTTCGCTGGCGCGTT 2917
 QY 363 AlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyrSerGly 382

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,35e-271 Length: 2699
 Score: 2950.00 Matches: 557
 Percent Similarity: 83.79% Conservative: 89
 Best Local Similarity: 72.24% Mismatches: 113
 Query Match: 67.03% Indels: 12
 DB: 4 Gaps: 7

US-10-028-384-2 (1-826) x ABL13247 (1-2699)

QY 63 SerGlnProAlaGlyTyrPrlGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
 DB 107 AGCAAGTGGCTGCTACAGAGCCTAATCACCTTCGCCATCTGCTAATCCCTGGGTG 166
 QY 83 AlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
 DB 167 GCGGATTTCTCTCGCCTCTTCGCGTCATCCGTTTCGAGTCGATTATCCATGAGTT 226
 QY 103 AspProTrpPheAntyArgSerThrHisIleLeuAlaSerHisGlyPheThrGluPhe 122
 DB 227 GATCCGTGGTTCAACTACCGGGCCACCGCTACATGTCGACAGATGTTGGTACAACTTC 286
 QY 123 LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyThrVal 142
 DB 287 CTCACCTGGTTTCGAGCGCGCATGATTCGCTCGGCAGGATGTTGGCGGTACCGTTC 346
 QY 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle 162
 DB 347 TATCCCGCCTGATGATTAGTCGCGGGAATCCATTTGCTGCTGACATCTCAACATA 406
 QY 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 DB 407 CCGGTCCATATTCGTGACATCTCGTGTCTTCGCGCCGATCTTCAGTGGCTGACCTCC 466
 QY 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 202
 DB 467 ATCTCCACCTACCTGCTGACCAAGAGCTGTGTCGCGGGCGCGCCCTCTTCGCGGCC 526
 QY 203 CysPheIleAlaIleValPrlCysGlyTyrIleSerArgSerValAlaGlySerPheAspAsn 222
 DB 527 AGCTTCATCGCATCGTGGCTGATCATCAGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 586
 QY 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleSerValIys 242
 DB 587 GAGGGCATTTGCCATTTTCCCTGCTGACATTCACCTGATCCCTGCGGTGGCTCAGTGAAG 646
 QY 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheThrMetValSerAla 262
 DB 647 ACTGGATCCGTTCTGCTGGCGCGAGCGCTTTGCTCTACTTCTACATGTTGTCGCGCC 706
 QY 263 TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeu 282
 DB 707 TGGGGTGGCTACGTTTTCATCATCACTGATACCCCTGACGCTCTTCGTATGCTGCTCAT 766
 QY 283 MetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
 DB 767 ATGGCAGGTACTCGCGCGTCTGCTGACCGAGCTACAGCACCTTCTACATCTCGGAGCTG 826
 QY 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 DB 827 CTGTTCTCCATGCGAGATCCCTCTCGTGGGATTCACCGGATACGACCGACGAGTGAACATG 886
 QY 323 AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
 DB 887 GCTGCGCTGGAGTGTGTGCTCTTATGGCCGTGGCCACCTTGGCCATTTGCGAGTCC 946
 QY 343 ArgLeuThrLysGlnGluPheGlnThrLysPheLeuGlyValSerLeuAlaGly 362
 DB 947 GTGCTGTCCGCAACAGAGTTCGGAAGCTGTTCATGTCGCGCGGATGCTGTGGTGGCGTT 1006

QY 363 AlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGly 382
 DB 1007 GCGGTCTTTTGGCGGTGCTGCTCACCATGCTGGCGGTGTTGGCGGTGGAGTGA 1066
 QY 383 ArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSer 402
 DB 1067 CGCTTCTACTCTGCTGGGATCTGGCTACGCCAAGATCCACATTCATCCATTCGATCC 1126
 QY 403 ValSerGluHisGlnProThrTrpValSerPhePhePheAspLeuHisIleLeuVal 422
 DB 1127 GTGTGCGAGCATCAGCCACCATCTGTTCTCTGTTCTCTTGTATCTGCACTCTCGTG 1186
 QY 423 CysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPhe 442
 DB 1187 TGGCGCTTCCAGTGGAGTGTGCTACTGCATCAAGCAGATCAACGAGAGCGGTTTC 1246
 QY 443 ValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 462
 DB 1247 GTGGTCTGTACGCCATCAGTGGGTCTTACTTCTGCTGTGTGTGTGTGTGTGTGTGTG 1306
 QY 463 ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482
 DB 1307 ACCCTCACCGCGTGTGTCATGCTGCGGAGTGGCTTTTCGGGACTGTTGGATGTG 1366
 QY 483 TyrLeuGlyAspAsp---MetLysArgGluAsnProValGluAspSerSerAspGlu 501
 DB 1367 TTCCTCAAGAGGATTCGCTTAAGCGAATGGGCGACAGCCATAAGCGGAGCCGCGAGTG 1426
 QY 502 Asp-----AspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArg 518
 DB 1427 GATGAAGCTCAGATTCATTCAGAGAGAGAGCGCTGTACGACAGGCTGGCAGCTG--- 1483
 QY 519 LysHisAlaThrGlnGlnLysThrGluGlu---GlyLeuGlyProAsnIleLysSer 537
 DB 1484 AAGCATCGTACTAAGCATGATGCCCCAGCAGGATGCTGGCGTCAGCTCCACCTCCAGAGT 1543
 QY 538 IleValThrMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
 DB 1544 ATGTATTTTGGCGGTCTAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1603
 QY 558 ThrSerAsnAlaTyrSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
 DB 1604 ACCAGCAATGCTTACTCCAGTCCCTCCATGCTTGTGGCTTTCCACAAAGTCAAGATGA 1663
 QY 577 ThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAsp 596
 DB 1664 TCCCGCAACATTTTAGACGATTCAGAGAGGCTTACTTACTGGCTTTCGACAGACACTGCC 1723
 QY 597 GluHisAlaArgValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 616
 DB 1724 GATGATGCTCGGCTATGCTTGTGGGATTTACGATACCCAGATAGCGGAATGGCAAC 1783
 QY 617 ArgThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 636
 DB 1784 AGAACCGACTAGTGGATTAATACCTGGAGAACATAGTACATAGCGCTGTTGGCAAG 1843
 QY 637 AlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyr 656
 DB 1844 GCAATGCTTCAACCGAGGAGAGTCTTACGAAATATGACATCTCTTGAAGTGGACTAC 1903
 QY 657 ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 676
 DB 1904 GTTTTGTGATCTTTGGCGGTGTGATCGCTATCTGCGGATGATATCAACAAGTTCCTG 1963
 QY 677 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPhe 696
 DB 1964 TGGATGCTCGGAATTTGCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTTACT 2023
 QY 697 ThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuAsnCysLeu 716
 DB 2024 ACCGACCGGCTGATTCAGGGTAGATGCCGAGGAGTGTCTCGGCCCTCTCACTGCTT 2083

QY 343 ArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaGly 362
Db 938 GTGCTGTCGCGCAACAGTTCGGAAGCTGTTCATCGTCGCGGATTCGTCGGCGGT 397
QY 363 AlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyrPheGly 382
Db 998 GCGGCTTTGTGCGCGTGTGCTCACCATGCTGGCGGTGTGGCCCGCTGGAGTGA 1057
QY 383 ArgPheTyrSerLeuThrTyrPheThrGlyTyrAlaLysIleHisLeuPheLeuVal 402
Db 1058 CGCTTCTACTGCTGCGGATCTGCTACCGCCAGATCCACATTCCTCATTCGATCC 1117
QY 403 ValSerGluHisGlnProThrThrTyrPhePhePhePhePheHisLeuVal 422
Db 1118 GTGTGGAGCATCAGCCACCATTTGTTCTGCTTCTTTCATCTGCACATCTCTGGT 1177
QY 423 CysThrPheProAlaGlyLeuThrPheCysIleLysAsnIleAsnAspGluArgValPhe 442
Db 1178 TCGGCTTCCCAAGTGGAGTGTGGTACGATCGATCAAGCAGATCAACGACGAGCGGTTC 1237
QY 443 ValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 462
Db 1238 GTGGTGTCTACGCCATCAGTGGGTTCATCTGCTGTGTGTGTGTGTGTGTGTGT 1297
QY 463 ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482
Db 1298 ACCCTCACCGCGT 1357
QY 483 TyrLeuGlyAspAsp---MetLysArgGluAsnProProValGluAspSerSerAspGlu 501
Db 1358 TTCCTGCAAGAGGATTCGTAAAGCGAATGGCCAGCAGCCATTAAGCGCAGCCAGGATG 1417
QY 502 Asp-----AspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArg 518
Db 1418 GATGAGCTGAGGATTCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1474
QY 519 LysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsnIleLysSer 537
Db 1475 AAGCATCGTACTAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1534
QY 538 IleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTyrPheVal 557
Db 1535 ATGTGTTATTTGGCGGTCTAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1594
QY 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
Db 1595 ACCAGCAATGCTACTCCAGTCCCTCCATTCCTGCTTTCACACAGCTCAAGATGA 1654
QY 577 ThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTyrLeuArgGlnAsnThrAsp 596
Db 1655 TCCCGCAACATTTAGACGATTCAGAGAGGCTTACTACTGCTTTCGACAGACATGCTCC 1714
QY 597 GluHisAlaArgValMetSerTyrTyrPheTyrGlyTyrGlnIleAlaGlyMetAlaAsn 616
Db 1715 GATGATGCTCGCGTATATGCTTGTGGGATTCAGGATACGATACGATAGCGGAATGGCAAC 1774
QY 617 ArgThrThrLeuValAspAsnAsnThrThrPheAsnAsnSerHisIleAlaLeuValGlyLys 636
Db 1775 AGAACGACCTAGTGGATAATAATACGCGAGACATAGTACATAGCGCTGTGTGGCAAG 1834
QY 637 AlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyr 556
Db 1835 GCAATGCTTTCACGAGAGAGAGGATTCACGAAATATGATCTCTGTGACGTGGACTAC 1894
QY 657 ValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 676
Db 1895 GTTTTGGTGATCTTTGGCGGTGTGTGCTATTCGCGATGATATCAACAAGTTCCTG 1954
QY 677 TrpMetValArgIleAlaGluGluHisProLysAspIleArgGluSerAspTyrPhe 696
Db 1955 TGAATGCTCGAATTCCTGAGGAGAGAGATCCCAAGGACATTAAGGAAACGCGATTACTTT 2014
QY 697 ThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeu 716

Db 2015 ACCGACCGCGTGAATTCAGGCTAGATCCGAGGTGCTCCGGCCCTGCTCAACTGCCTT 2074
QY 717 MetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProPro 736
Db 2075 ATGTACAAATTAAGCTACTACAGATTTCGGGAATTTGAAGTTGGACTACAGAGGTCCATCT 2134
QY 737 GlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeu 756
Db 2135 GGATATGATCGCACAGTAACCGCTCATTTGGGAATAAGGACTTCGATCTGACCTACTCTG 2194
QY 757 GluGluAlaPheThrSerGluHisTyrPheLeuAlaArgIleTyrLysValLysAlaProAsp 776
Db 2195 GAGGAGCGCTACACACAGACACTGCTGTTCGCTATCTATAGGTTGAAGGACCGCAT 2254
QY 777 -----AsnArgGlnThrLeuAspHisLysProAspValThrAsnIlePheProLysGln 794
Db 2255 GAGTTCTATAGACCATCTCTGAAGCAAGAGAGAGAG-----ATTCTCCAGCA 2305
QY 795 LysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuVal 814
Db 2306 AACTTTCATTTCCGAGAAAGAACTCTAAGCGTCGCAAGGCTCATACAGAAACCGCGGT 2365
QY 815 PheLysLysGlyLysLysLysLysSerLysLysThr 825
Db 2366 GTTGTAAAGGAAACGAACTTGAAATAAAC 2398
RESULT 4
ABL13247
ID ABL13247 standard; cDNA; 2699 BP.
XX ABL13247;
AC ABL13247;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li FWD, Myers EW;
PI WPI; 2001-656860/75.
DR P-PSDB; ABB69144.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 34223; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly

Qy 760 PheThrSerGluHieTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGlu 779
 Db 2345 TTTACATCAGACCACTGGCTTGTGAGGATATATAAGTGAAGAACACCTGACACAGGGAG 2404
 Qy 780 ThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLys 799
 Db 2405 ACACTAGGTCAACAACCTCGAGTCAACAATCGTCCCAACAGAGATTTGTCAAG 2464
 Qy 800 LysThrThrLysArgLysArgGlyTyrIleLysLysLysLeuValPheLysLysGlyLys 819
 Db 2465 AAGACTACTAAAGGAAGCGTGGCTACGTTAAAAATAAGTAGTGTGTTAAGAAAGCAAG 2524
 Qy 820 LysIleSerLysLysThrVal 826
 Db 2525 AAGACCTCTAAGAAAGACTGTT 2545

RESULT 3

ADD94789
 ID ADD94789 standard; DNA; 2417 BP.

XX AC ADD94789;

XX 29-JAN-2004 (first entry)

XX Drosophila melanogaster STT3 gene sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytostatic; immunosuppressive; antisense therapy; Gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide;
 KW fruit fly; gene; ds; STT3.

XX Drosophila melanogaster.

XX WO2003054008-A2.

XX 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

XX 20-DEC-2001; 2001US-00028384.

XX (COMP-) COMPATIGENE INC.

XX Perreault C, McBride K;

XX WPI; 2003-559122/52.

XX P-PSDB; ADD94790.

XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.

XX Claim 6; SEQ ID NO 7; 66pp; English.

XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response

CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the Drosophila melanogaster STT3 gene which
 CC is related to the invention. Note: This sequence does not appear in the
 CC specification but was obtained by the indexer from GenBank.

SQ Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.14e-271 Length: 2417
 Score: 2950.00 Matches: 557
 Percent Similarity: 83.79% Conservative: 89
 Best Local Similarity: 72.24% Mismatches: 113
 Query Match: 67.03% Indels: 12
 DB: 10 Gaps: 7

US-10-028-384-2 (1-826) x ADD94789 (1-2417)

Qy 63 SerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
 Db 98 AGCAAGGTGGCTGGCTACAGACGCTAATCACCTTGGCCATCTGCTTAATCGCTGGCTG 157
 Qy 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHisGluPhe 102
 Db 158 GCCGATTTCCTCTCGCCTCTTCGCGTATCTGTTTCGAGTCGATATCCAGAGTTT 217
 Qy 103 AspProTrpPheAsnTyrArgSerThrHisLeuAlaSerHisGlyPheTyrGluPhe 122
 Db 218 GATCCCGTGTCAACTACCGGCCACCGCTACATGGTGCAGATGGTGTGTACAACTTC 277
 Qy 123 LeuAsnTrpPheAsnGluArgAlaTrpTyrProLeuGlyArgIleValGlyThrVal 142
 Db 278 CTCACTGTTTCAGACAGCGCATGGATCCGCTCGCAGCATGTGGCGGTACCGTC 337
 Qy 143 TyrProGlyIleuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle 162
 Db 338 TATCCCGGCTGATGATTACGTCGCGGGAATCCATTGGCTGTGTGCAGTACTCAACATA 397
 Qy 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 Db 398 CCGGTCCATATCGTGACATCGCTGTCTCGCGCGGATCTTCAGTGGCTGACCTCC 457
 Qy 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 202
 Db 458 ATCTCCACCTACCTGCTGACCAAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 517
 Qy 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn 222
 Db 518 AGCTTCATGCGCATGCTGCTGGCTACATCATAGTAGTGGTGGTGGTGGTGGTGGTGG 577
 Qy 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLys 242
 Db 578 GAGGGGATTCGATATTCGCTGAGTTCACCTTCTCTGTGGTGGTGGTGGTGGTGGTGG 637
 Qy 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
 Db 638 ACTGGATCCGTGTTCTGTGTGGCCGACCGCTTGTCTTCTTCTTCTTCTTCTTCTTCT 697
 Qy 263 TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeu 282
 Db 698 TGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 757
 Qy 283 MetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
 Db 758 ATGGGAGGTACTTCGCGCGTCTGTGTGACCGACCTACAGCACCTTCTTACATCTTGGG 817
 Qy 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 Db 818 CTGTTCTCCATGAGATCCCTTCTGTGGGATTCGATACGATACGATACGATACGATACG 877
 Qy 323 AlaAlaIleAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuAsp 342
 Db 878 GCTGCGCTGGAGTGTGTTGTGTCCTTATGGCCGTGGCCACCTTGGCCCATTTTGCAGTC 937

Db 137 AGCGGCTCATGGCTCTGGGGAACAGCGCGCACCGGACCATGGGCGCGGAACCCAGAGC 136
Qy 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAla---ProAlaGlyLeuSer 59
Db 197 GCGTCCAGGCGCGCGCG-----CGGAAGCGCGCGCGCGCGCGCGCGCGCGTGC 244
Qy 60 GlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeu 79
Db 245 GCGGGCTTGTGGAGCGCGCGCGGGTGGAGTGTGTCTCTCTTCACCAATCCCTTCTTG 304
Qy 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIle 99
Db 305 GCGTGGCTGGCGGCTTCAGCTCGCGGCTCTTCGCGGTCATCCGCTTCGAGAGCATATC 364
Qy 100 HisGluPheAspProTrpPheAsnTrpArgSerThrHisHisLeuAlaSerHisGlyPhe 119
Db 365 CAGGAGTTCGACCGCGGTTCACCTATAGATCAACACATCATCTTGATCTCATGATTC 424
Qy 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGly 139
Db 425 TATGAGTTCCTAAATGGTTTCATGAAGAGCATGGTACCACCTGGGAAGATAGTGGGT 494
Qy 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
Db 485 GGCACCGTTTACCAGGGTTCATGATTAACAGCTGGCCCTTATTCATGGATTTTAAATACA 544
Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
Db 545 TTGAACATACAGTTCACATAGAGATGTGTGTPATTCTCTTGACCAACTTTTAGCGGC 604
Qy 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199
Db 605 CTTTACATCCATATCTACGTTCTCTGCTAACTAGAGAACTGTGGAACCAAGAGCAGGACTT 664
Qy 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
Db 665 CTAGCTGCTGCTTCATCTGCTATCGTACCAGGATACATATCTCGCTCAGTGGCGGATCC 724
Qy 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLys 239
Db 725 TTTGATAATAGAGCAATGCCATTTTTCGGCTTCAGTTCACTTACTTATGGGTAAAG 784
Qy 240 SerValTyrThrCysSerValPheThrMetCysCysValSerTyrPheThrMet 259
Db 785 TCTGTGAAGCCGGTCTGTGTCTGACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 844
Qy 260 ValSerAlaTrpGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
Db 845 GTCTCTGCGTGGGAGGTATGTGTTCATCATCAACCTCATCCCTCTCCATGTTGTGTG 904
Qy 280 LeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIle 299
Db 905 TTGCTGCTGATGCAGAGGTACAGCAAGAGTCTACATAGCATATAGCATTTGTACATT 964
Qy 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
Db 965 GTGGGTATTAATATATCCATGAGATACCTTTTGTGGGATTTACGCCAATCAGAACAAAGC 1024
Qy 320 GluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
Db 1025 GAGCATATGGCAGCTCAGGTCTCTTGGCGTGTGCAAGCTTACGCTTTTTCAGATAT 1084
Qy 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
Db 1085 CTGAGAGACCGGTGCAAAACAGGAGTTCCAGACCCCTTTTCTTTTGGGTGTCTCACTA 1144
Qy 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
Db 1145 GCTGCGGCGCTGTCTTCTAGTGTCTATCTGACATACACAGGTTATATGACCA 1204
Qy 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399
Db 1205 TGGAGTGGCAGGTTTTTATTCATATGGGATATCGGATGTCGGAATAACACATTCACATT 1264

Qy 400 IleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePhePheAspLeuHis 419
Db 1265 ATTGCATCAGTGTCTGAACATCAGCCTAGCATGGGTGTCTTTCTTTTGTATCATAT 1324
Qy 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLeuAsnIleAsnAspGlu 439
Db 1325 ATTCTTGTATGTACCTTCCAGCAGGCTTATGGTCTGCATCAAAAATATCAACGATGAA 1384
Qy 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
Db 1385 AGAGTATTTGTGCTCTGTATGCGATCAGTGTGTGTACTTTGCGCGAGTGTATGTGCGG 1444
Qy 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
Db 1445 CTGATGCTGACTCGACCCCGGTCTGCTGCATGCTGTGGCCATCGCTTCTCCAATGTT 1504
Qy 480 PheGluHisTyrLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSer 499
Db 1505 TTTGAGCAGTATTTGGGGGATGACATGAAGGGGAAACCCACCTGTGGAGACAGCAGT 1564
Qy 500 AspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLys 519
Db 1565 GATGAGGATGACAAAGAAACCCAGGAAACTTGTATGACAAGGCAAGTAAAGTAGGAGAG 1624
Qy 520 HisAlaThrGlnGlnGlnLysThrGluGluGlyLeuGlyProAsnIleLysSerIleVal 539
Db 1625 CATGTGACAGCAAGAGAAACCTGAAGAGGGCTTGGGCCCCCAACATCAAAAGCAATTGTG 1684
Qy 540 ThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSer 559
Db 1685 ACCATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1744
Qy 560 AsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsn 579
Db 1745 AACCCCTACTCCAGTCCAGTGGTCTTGGCTCTCTACATCATGTGTGTGTGTGTGTGTGT 1804
Qy 580 IleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAla 599
Db 1805 ATATTAGATGATTTAGAGAAAGGTACTTTTGGCTGAGACAAACACGATGAAACACGCC 1864
Qy 600 ArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThr 619
Db 1865 CGGCTCATGTCTGTGTGGGACTACGCTATCAGATTGTGGCATGGCCACACGACCACT 1924
Qy 620 LeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSer 639
Db 1925 CTGTTGGATACAAACCTGGAAACACAGCCACATCGCACTGCTGCGGAAAGCTATGTCT 1984
Qy 640 SerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuVal 659
Db 1985 TCCATGGAACCGCGCTATATAAATCATGAGGTCCCTTGATGTGATATGTGTGTGTGT 2044
Qy 660 IlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetVal 679
Db 2045 ATTTTCGAGAGTGTATGCTATTCCGGGAGCATATCAACAGTTCTCTCTGGATGTC 2104
Qy 680 ArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGln 699
Db 2105 AGGATAGCTGAAGGGAGCATCCCAAGACATCCCGGAAGGTGACTATTTTACCACGAG 2164
Qy 700 GlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLys 719
Db 2165 GGAGAGTTCGAGTAGCAAAAGCTGGTCTCTACTCTCTTAACTGCTTATGTATATAA 2224
Qy 720 MetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAsp 739
Db 2225 ATGTCATACATACAGTTTGGAGAAATGACGTAGATTTTCGCACATCCCGGAGCTTTCAC 2284
Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAla 759
Db 2285 CGAACACGTAATCTGAGATTGGAATAAAGACATTAATAATTCAAGCATTTGGAGGAGCT 2344

Db 1501 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGCGAGTAAAGTGAGGAACAT 1560
 Qy 521 AlaThrGluGlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleValThr 540
 Db 1561 GCNACTGACAGGAAACAACTGAAGAGGATAGGCCCTAATAATAAAGACATGTGCACC 1620
 Qy 541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
 Db 1621 ATGTTGATGCTGATGCTATGATGATGTTGCTGCTCCTACTGCTGCTGCTGCTGCTGCT 1680
 Qy 561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
 Db 1681 GCCTACTCTAGTCCAGAGTACTGCTGCTCCTCATACATCATGATGCGACACAGGATATC 1740
 Qy 581 LeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
 Db 1741 TTAGATGATGTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACGA 1800
 Qy 601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
 Db 1801 GTAATGCTTGGTGGGATATGGCTATCAGATAGCTGGATGGCTATAGACTAGCTTGG 1860
 Qy 621 ValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
 Db 1861 GTGGATAATAACACCTGGAAATACACGACACATAGCTGTGGGAAAGCTATGTCTTCT 1920
 Qy 641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIle 660
 Db 1921 AATGAACAGCAGCCTATAAATCATGAGGACTCTAGATGTAGATTATGTTTGGTTAT 1980
 Qy 661 PheGlyGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrpMetValArg 680
 Db 1981 TTTGGAGGGGTATTTGGCTATTCTGCTGATGATATCAACAATTTCTCTGGATGGTAGG 2040
 Qy 681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
 Db 2041 ATAGCTGAAGGAGAAACATCCAAACACATTCGGGAAGTGACTATTATCCCCACAGGGA 2100
 Qy 701 GluPheArgValAspIleAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
 Db 2101 GAATTCGGGTGAGCAAGACAGGATCCCTACTTGTGTGATTCGCTTATGTATAAAATG 2160
 Qy 721 SerTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArg 740
 Db 2161 TCATACTACAGATTGGAGAAATGCAGCTGGATTTCGTACACCCCGAGGTTTGCACCGA 2220
 Qy 741 ThrArgAsnAlaGluIleGlyAsnLysPheLysPheLysHisLeuGluAlaPhe 760
 Db 2221 ACACGTAATGCTGAGATTGGAATTAAGGACATTAATTCAAACATTTCGGAAGAGCCTTT 2280
 Qy 761 ThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
 Db 2281 ACATCAGACACTGGCTGTTAGGATATATAAGTAAAGCACCTGATACAGGAGACAGA 2340
 Qy 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
 Db 2341 TTAGATCACAACCTTCAGTGCACCAACATTTTCCAAACACAGAAATTTTGTCAAGAG 2400
 Qy 801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
 Db 2401 ACTACCAAGAGGAGCGTGGCTACATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAA 2460
 Qy 821 IleSerLysLysThrVal 826
 Db 2461 ATATCTAAGACACTGTT 2478

RESULT 2

ADD94785

ID ADD94785 standard; cDNA; 2710 BP.

XX XX

AC ADD94785;

XX XX

DT 29-JAN-2004 (first entry)

XX Mouse SIMP cDNA sequence.
 DE source of immunodominant MHC-associated peptide; SIMP; MHC;
 XX major histocompatibility complex; human leukocyte antigen; HLA;
 KW cyclostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 KW murine; gene; ss.
 OS Mus musculus.
 XX WO2003054008-A2.
 PN 03-JUL-2003.
 XX 18-DEC-2002; 2002WO-CA001967.
 XX 20-DEC-2001; 2001US-00028384.
 XX (COMP-) COMPATIGENE INC.
 XX Perreault C, McBride K;
 DR WPI; 2003-559122/52.
 DR P-PSDB; ADD94786.
 XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 PS Disclosure; SEQ ID NO 3; 66pp; English.
 CC This invention relates to a novel isolated or purified human protein, termed
 CC source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the mouse SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX Sequence 2710 BP; 678 A; 648 C; 689 G; 695 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2710
 Score: 4263.50 Matches: 805
 Percent Similarity: 97.70% Conservative: 3
 Best Local Similarity: 97.34% Mismatches: 14
 Query Match: 96.88% Indels: 5
 DB: 10 Gaps: 2

US-10-028-384-2 (1-826) x ADD94785 (1-2710)

Qy 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerProTrp 20

Db 77 ATGGCGAGCGCTTCGGCCCCCGAGAGCAAGCAAGTCGCTCCCTCAACTGCTCCCGTGG 136

Qy 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40

XX PI Perreault C, McBride K;
 XX DR WPI; 2003-559122/52.
 XX DR P-PSDB; ADD94784.
 XX PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX PT
 XX PT
 XX PS Claim 6; SEQ ID NO 1; 65pp; English.
 XX CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX CC
 XX SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 2481
 Score: 4401.00 Matches: 826
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0
 US-10-028-384-2 (1-826) x ADD94783 (1-2481)
 QY 1 MetAlaGluProSerAlaProGluSerIleHisGlySerSerLeuAsnSerSerProTrp 20
 DB 1 ATGGCGGAGCCCTCGGCCCGCGAGCAAGCAAGTCGTCCTCAACTCGTCCCGTG 60
 QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 DB 61 AGTGGCCCTCATGGCCCTGGGAACAGCGCGGACGCGCCACCGCGGCCGCCAGTCG 120
 QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 DB 121 GGCACAAAGCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
 QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 DB 181 GGCCTGTGCGAGCG 240
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis 100
 DB 241 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAAAGCATATCCAC 300
 QY 101 GluPheAspProTrpPheAsnTyArgSerThrHisIleuAlaSerHisGlyPheTy 120
 DB 301 GAGTTCGACCCCGTGTAACTATAGATCAACACATCATCTTCCTTCACCACTCTCAT 360
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTyProLeuGlyArgIleValGlyGly 140
 DB 361 GAATTTTAAATTCGTTTATGTAAGAGCATGGTATCCATCAGGAAGATAGTAGTGT 420
 QY 141 ThrValTyProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160

DB 421 ACTGTTTACCAGCGTTGATGATAACCGCTGCGCTTATTTCATTGGATTAAATACATTG 480
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 DB 481 AACATAACTGTTTACATATGAGAGAGCGTATGTGTGTTCTTGCACCAACTTTTAGCGGCTT 540
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
 DB 541 ACATCTATATCTACTTCTCTGCTTACAGAGAACCTTTGGAAACCAAGGAGGAGGACTTTA 600
 QY 201 AlaAlaCysPheIleAlaIleValProGlyTyIleSerArgSerValAlaGlySerPhe 220
 DB 601 GCTGCTGTTTATTCCTATTGACAGGCTACATATCTCGCTCAGTAGCTGATCTT 660
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyTyIleuTrpVallysser 240
 DB 661 GATAAATGAAGGCAATGCTATTTTTCACCTTCAGTTTCACATCTATTTATGGGTAAATCT 720
 QY 241 ValIysThrGlySerValPheThrMetCysCysValSerCysLeuSerTyPheTyMetVal 260
 DB 721 GTAAAACTCGGTCAGTTTGGACAATGCTGCTGCTTATCTATCTATATGCTC 780
 QY 261 SerAlaTrpGlyGlyTyValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
 DB 781 TCTGCTTGGGCTGGTATGTTTATCATCAATCTTATTCCTCCTCATGTATTTGTGTG 840
 QY 281 LeuLeuMetGlnArgTySerLysArgValTyIleAlaTySerThrPheTyIleVal 300
 DB 841 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACCTTCTTACATTGTG 900
 QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 DB 901 GGTTTAAATATCAATGAGATACCTTTTGGGAATCCAGCAATCAGAACAGTGA 960
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyAlaPheLeuGlnTyLeu 340
 DB 961 CACATGSCAGCTCGAGGTGCTTTGTCATTTGTCAGCTTATGCTTCTTTCAGTATCTG 1020
 QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 DB 1021 AGAGACGATTAACAAACAAAGAGTTCCAGAGCCCTTTCTTTTGGGTATACATAGCT 1080
 QY 361 AlaGlyAlaValPheLeuSerValIleTyLeuThrTyThrGlyTyIleAlaProTrp 380
 DB 1081 GCAGTCTGCTGTTCTTCTAGTGTCACTATTGACTTATACAGGTTACATTGCACCAATG 1140
 QY 381 SerGlyArgPheTySerLeuTrpAspThrGlyTyAlaLysIleHisIleProIleIle 400
 DB 1141 AGTGGCAGGTTTTATTCTATTTGGGATACCTGGGTATGCAAAATACACATTCCCAATTAT 1200
 QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePhePheAspLeuHisIle 420
 DB 1201 GCATCAGTGTCTGAGCATCAACCTAGACTTGGGTGCTTTCTTCTTGTATCATATT 1260
 QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
 DB 1361 CTGTATGTACCTTCCAGCAGGCGCTTTGGTCTGCTATGCAAAATATCAACGATGAAAGA 1320
 QY 441 ValPheValAlaLeuTyAlaIleSerAlaValTyPheAlaGlyValMetValArgLeu 460
 DB 1321 GTATTTGCTCTATATGCAATCAGTGTGCTGCTTCTTCTTGTGGAGTGTGTCGCGCTG 1380
 QY 461 MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
 DB 1381 ATGTGTACTTTGACTCCAGTCGTGTATGCTGTCTGCAATTTGCTTTTCAAAATGTTTT 1440
 QY 481 GluHisTyLeuGlyAspMetLysArgGluAsnProProValGluAspSerSerAsp 500
 DB 1441 GAGCACTATTTCGGGATGACATGAAGGGAATCCACCTGTGCGAGCAGCAGTGTAT 1500
 QY 501 GluAspAspLysArgAsnGlnGlyAsnLeuTyAspLysAlaGlyValArgLysHis 520

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:28:05 : Search time 951.963 Seconds
(without alignments)
4554.822 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 4401

Sequence: 1 MAEPSAPESKHKSLSNPPW.....GVKNKLVFKGKISKTKTV 826

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Xgapop 10.0			
Xgapext 0.5			
Fgapop 6.0			
Fgapext 7.0			
Delop 6.0			
Delext 7.0			

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-Q/cgn2.1/USPTO_spoil/US10028384/runat_14122004_131518_8007/app_query.fasta_1.3740
-DB=N Geneseq_23Sep04 -QMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTWT=prc -NOGM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 -CGN_1_1396 -runat_14122004_131518_8007 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPOP=0.5 -DELOP=6 -DELEX=7

Database : N Geneseq_23Sep04:*

	1: geneseqn1980s:*
2:	geneseqn1990s:*
3:	geneseqn2000s:*
4:	geneseqn2001as:*
5:	geneseqn2001bs:*
6:	geneseqn2002as:*
7:	geneseqn2002bs:*
8:	geneseqn2003as:*
9:	geneseqn2003bs:*
10:	geneseqn2003cs:*
11:	geneseqn2003ds:*
12:	geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4401	100.0	2481	10	ADD94783 Human SIM
2	4263.5	96.9	2710	10	ADD94785 Mouse SIM
3	2950	67.0	2417	10	ADD94789 Drosophila
4	2950	67.0	2699	4	ABL13247
5	2838	64.5	4922	4	ABL13246
6	2370	53.9	1664	4	AAK94164 Human ful

7	2370	53.9	1664	12	ADL130661
8	2274	51.7	2855	4	ABL02795
9	2266.5	51.5	3094	10	ADD94791
10	2250	51.1	2472	10	ADD94793
11	2250	51.1	2472	10	ADH28844
12	2236	50.8	2760	5	ABV24502
13	2181.5	49.6	2466	10	ADB69900
14	2179.5	49.5	2332	8	ABT20789
15	2164	49.2	2356	6	ABZ32036
16	2023	46.0	2803	8	ABT20181
17	2023	46.0	3969	8	ABT17781
18	2023	46.0	4603	8	ABT19595
19	2021.5	45.9	2733	10	ADD94787
20	2021.5	45.9	6153	4	ABL02794
21	1958.5	44.5	2000	10	ACC61293
22	1958.5	44.5	2000	10	ADK63659
23	1941.5	44.1	4738	10	ADK63659
24	1939.5	44.1	2882	10	ADK63659
25	1897	43.1	1848	8	ABT18969
26	1896.5	43.1	1543	2	AAK85055
27	1896.5	43.1	1543	8	ACD18981
28	1896.5	43.1	1543	12	ADG78372
29	1896.5	43.1	1543	12	ADN60663
30	1894.5	43.0	1209	4	AAH33264
31	1894.5	43.0	1209	6	ABL89850
32	1855	42.1	1369	8	ABT18375
33	1722	39.1	2546	2	AAV44866
34	1722	39.1	2546	5	AAF98463
35	1689	38.4	2510	5	AAF93772
36	1586	36.0	2537	4	AAO08289
37	1581	35.9	2547	4	AAO08315
38	1581	35.9	2660	6	ABO54750
39	1552.5	35.3	1114	4	AAH99794
40	1531	34.8	2384	4	AAH18021
41	1370.5	31.1	1371	12	ADP28508
42	1253	28.5	787	4	AAH07526
43	1200	27.3	764	5	AAF93968
44	907	20.6	507	3	AAO01967
45	872	19.8	2953	4	ABL18224

ALIGNMENTS

RESULT 1

ADD94783
ID ADD94783 standard; cDNA; 2481 BP.

XX

AC ADD94783;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human SIMP cDNA sequence.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

AFLOYLRDLTKQBFOTLFLGLVSLAAGAVFLSVLYLTGYIAPWSGRFYSLMDTGY
AKIHPIIASVSEHQPTTWVSFFDLHLVCTPPAGLWFCIKINIDRVFGA"

ORIGIN

Alignment Scores:
Prod. No.: 5, 94e-191 Length: 1664
Score: 2370.00 Matches: 443
Percent Similarity: 99.77% Conservatives: 1
Best Local Similarity: 99.77% Mismatches: 1
Query Match: 53.85% Indels: 0
DB: 6 Gaps: 0

US-10-028-384-2 (1-826) x CQ782554 (1-1664)

QY 1 MetalAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTyr 20
Db 28 ATGGCGAGCCCTCGGCCCGGAGAGCAGCAGCGTCCCTCAACTCGTCCCCGTGG 87
QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
Db 88 AGTGGCCTCATGGCCCTGGGAAACAGCGCCAGCCACCGCGCGCCGCCAGTGC 147
QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
Db 148 GGGCACRAGGCG 207
QY 61 GlyLeuSerGlnProAlaGlyTyrGlnSerLeuLeuSerPheThrLeuPheLeuAla 80
Db 208 GGGCTGTCCGACCGCGCGTGGGTGGCAGTGGCTTCTCTCCATCCATCTCTCTCC 267
QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis 100
Db 268 TGGCTTGGCGGCTTCAGCTCGGCTCTTCCGCGTCTTCCGCGTCTTCCGCGCT 327
QY 101 GluPheAspProTyrPheAsnTyrArgSerThrHisLysLeuAlaSerHisGlyPheTyr 120
Db 328 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTTGCATCTCATGGGTTCAT 387
QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTyrTrpProLeuGlyArgIleValGlyGly 140
Db 388 GAATTTTAAATGGTTTGGTAAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGT 447
QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisThrIleLeuAsnThrLeu 160
Db 448 ACTGTTTACCCAGGCTTATGATACCGCTGCTTATTCATGGATTTTAAATACATTG 507
QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
Db 508 AACATAACTGTTTACATAAGAGACGTGTGTGTCTTCTGCACCAACTTTTAGCGCCT 567
QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
Db 568 ACATCTATATCTACTTCTCTTACAAAGAACTTTGGAAACCAAGAGGAGGACTTTTA 627
QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
Db 628 GCTGCTTGTATTATGTCTATTACAGGCTTACATATCTCGGTACGTAGTGGATCCTTT 687
QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTrpLeuTrpValLysSer 240
Db 688 GATATAGAGGCTTGTCTATTTCACCTTCAGTTCACATCTATTATGGGTAATCT 747
QY 241 ValLysThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetVal 260
Db 748 GTAAAACTGGGCTAGTTTGGACAATGTGCTGCTTATCTTATTTCTATATGCTC 807
QY 261 SerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
Db 808 TCTGCTTGGGCTGTATGTATTATCAATCAATCTTATTCACGTGATGTTTGTGTG 867
QY 281 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
Db 868 TTACTGATCAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTGTG 927

QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
Db 928 GGTTTAATATTATCAATGCAGATACCTTTTGTGGGATTCACGCCAATCAGAACAGTGAA 987
QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
Db 988 CACATGGCAGCTGCAGGTGTCTTTCATGTGTGCNAGCTTATGCTTTCTTGCAGTATCTG 1047
QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
Db 1048 AGAGACCGATTAAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGTATCACTAGCT 1107
QY 361 AlaGlyValAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyr 380
Db 1108 GCAGGTGCTGTCTTCTTGTCTCATCTATTGACTTATACAGTTACATTGCACCATGG 1167
QY 381 SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIle 400
Db 1168 AGTGGCAGGTTTTATTTCATTTGGGATACCTGGGTATGCAAAAATACACATTCCCAATTAT 1227
QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIle 420
Db 1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGTCTTTCTTCTTGTATCTACATATT 1287
QY 421 LeuValCysThrPheProIleGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
Db 1288 CTTGTATGTACCTTCCACAGCGCCTTTGGTTCGTCAATCAAAAATATACAGCATGAAGA 1347
QY 441 ValPheValAla 444
Db 1348 GTATTTGGTGTCT 1359

Search completed: December 15, 2004, 05:31:26
Job time : 9979.01 secs


```

/db_xref="wormBase:T12A2.15b"
/translation="MRKKWTIVLCITLQATFSLYHDSERKSLMERPKSHKCFSCA
SFVILPWSQMHYPPKFTDRCPQSGIGLVCSSACFTLVERDGEYVSEQH
GVIRGCDRLLLGLDDVRNIIISAYENQRVCHTRKLLRFLPLSGQDVVTFPCSN
GDFNEHDMRLSELSSNFQIHSVLLIVATWILIF"
join(3241..3365,3586..3740,6392..6558,6795..6926)
/gene="hot-4"
/standard_name="T12A2.15a"
/notes="coded for by the following C. elegans cDNAs:
AF324056"
/codon_start=1
/product="Homolog of odr-2(two) protein 4, isoform a"
/protein_id="AAM15600.1"
/db_xref="GI:20198878"
/db_xref="wormBase:T12A2.15a"
/translation="MRKKWTIVLCITLQATFSLYHDSERKSLMERPKSHKCFSCA
SFVILPWSQMHYPPKFTDRCPQSGIGLVCSSACFTLVERDGEYVSEQH
GVIRGCDRLLLGLDDVRNIIISAYENQRVCHTRKLLRFLPLSGQDVVTFPCSN
GDFNEHDMRLSELSSNFQIHSVLLIVATWILIF"
complement(7287..12051)
/gene="T12A2.15"
complement(join(7287..7364,7412..7520,7644..7735,
7844..7979,8026..8132,9511..9663,9726..9884,9932..10111,
10156..10458,10598..10687,10735..10911,10953..11156,
11304..11499,11832..11919,11982..12051))
/gene="T12A2.15"
/standard_name="T12A2.15a"
/notes="contains similarity to Pfam domain PF00168 (C2
domain); coded for by the following C. elegans cDNAs:
yk194a9.3, yk194a9.5, yk195a9.5, yk181e11.3, yk181e11.5"
/codon_start=1
/product="Hypothetical protein T12A2.15a"
/protein_id="AAC24452.2"
/db_xref="GI:20198877"
/db_xref="wormBase:T12A2.15a"
/translation="MSQSLVPLVGSAILSTFPFLGKWDYSFWVLLIIIASVTKS
FLYRRRELIARATARERIVMAQLQDLPAWQFPDTERVEWLNKVIHQLPYVG
EYTKFVNDRIIPQVAKQMGFKNFKFTMDMDGIPCRVGIGKVTNNVGRDLIVD
MDVAYAGDAFTFSCCGTGMNIIQFSGKRLAIKPLPVPVPMVGVSGTFLPMKPG
DNLTGMEVVELPGLDARSINSQIAALCVLNEIIVPLAPDVVTVLYFPBPDG
VRLKIEAKNLRDISPTFKGKSDPYAEIOVGSOFFKTRTIDDLNIPNEXPEAV
VDOAGDKLLELDFEDQKDESLRSLVDLKLVOAKGTIDKWPYLEGKGGDLHIKA
TWNLSLEHLEKQEWAEWQADPIHSAALLMYIDSVADLPFKSLPSPFVEV
SLGKETQTPVKTVNPLPQSKPLFFVRHLEQELKFEAVDDGTRRSLSNIPFLT
LLKEPNLEQNMHMLTGLVQSPFVITTRIMRSDSQSLNSHGRNSRLNIPFLT
KHENKETRADENRGEIEIQIDFDLNLQKIALIRCDLMTFPKQCCNPYGSVKL
VALDGNKEVFKKTPAKTRPHFDNVEIDINPDSLLNKHVINXVDNINYGTFVA
KPLVGLCLEIRLDSLMNRQLSQRWIPLSVERK"
complement(join(7287..7364,7412..7520,7644..7735,
7844..7979,8026..8132,8386..8532))
/gene="T12A2.15"
/standard_name="T12A2.15b"
/notes="contains similarity to Pfam domain PF00168 (C2
domain); coded for by the following C. elegans cDNAs:
yk349e10.5"
/codon_start=1
/product="Hypothetical protein T12A2.15b"
/protein_id="AAC24451.1"
/db_xref="GI:3238591"
/db_xref="wormBase:T12A2.15b"
/translation="MNELSAPFIERANGKVGSDVVAEPKTVNQDADDEVEIKLN
VDNQMLRSQSLSHSGNSRLGRILFRSHEMKKRETRADENRGEIEIQIDFDL
VNQKIALIRCDLMTFPKQCCNPYGSVKLVALDGNKEVFKKTPAKTRPHFDN
HVEIDINPDSLLNKHVINXVDNINYGTFVAKPVLGCLLEIRLDSLMNRQLSQRWIPLS
VERK"
complement(16387..22005)
/gene="T12A2.5"
complement(join(16387..16422,16786..17621,17892..18452,
18948..19052,19995..20008,20190..20562,20859..20951,
21039..22005))
/gene="T12A2.5"
/standard_name="T12A2.5"
/notes="contains similarity to Oryza sativa Putative myosin

```

heavy chain-like protein.; TR:Q94DC2; coded for by the following C. elegans cDNAs: CEBS524F"

Alignment Scores:

Pred. No.:	1-02e-197	Length:	54118
Score:	2605.00	Matches:	502
Percent Similarity:	70.61%	Conservative:	101
Best Local Similarity:	58.78%	Mismatches:	139
Query Match:	59.19%	Indels:	113
DB:	.3	Gaps:	8

US-10-028-384-2 (1-826) x U13019 (1-54118)

Qy	63	SerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTriPLeu	82
Db	31346	AGTAGTCGGCTCGAGCGACACTCTTCTTACCATCGTACTGGCTCGGCTGGCATGTTTC	31405
Qy <td>83</td> <td>AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe</td> <td>102</td>	83	AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe	102
Db	31406	GTCGGATTGGCTCCCGCCCTTTTCGCCATCGCCCTTGTAGTCGATTATCCACGAGTTC	31465
Qy <td>103</td> <td>AspProTrpPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyrGluPhe</td> <td>122</td>	103	AspProTrpPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyrGluPhe	122
Db	31466	GATCCATGGTTAACTATATCGGCTACACATCATATGGTTCAACATGGTTCTATAAATTC	31525
Qy <td>123</td> <td>LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyThrVal</td> <td>142</td>	123	LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyThrVal	142
Db	31526	TTGAAATGGTTTATGAGCGCTGGTATCCACATGGTGGTATGTCGGAGGAACTGTA	31585
Qy <td>143</td> <td>TyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsnThrLeuAsnIle</td> <td>162</td>	143	TyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsnThrLeuAsnIle	162
Db	31586	TATCCAGGACTCATGTGTAACATCTGGATTAATCTACTGGATTTCCTTAACATTC	31645
Qy <td>163</td> <td>ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer</td> <td>182</td>	163	ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer	182
Db	31646	CATGTTCAACATCGTGAAGTTTCGGTATTTCTGGCTCCAAACATCTCTGGCTGACTGCC	31705
Qy <td>183</td> <td>IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAsnGlyLeuLeuAla</td> <td>202</td>	183	IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAsnGlyLeuLeuAla	202
Db	31706	ATTGCCACCTATCTACTTAAAGGAGTTGTGGAGCCGAGGAGCTGGTCTTTTCCGCT	31765
Qy <td>203</td> <td>CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAsn</td> <td>222</td>	203	CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAsn	222
Db	31766	TGCTTTCATGCTATTTCTCCAGGATATACCTCCCTTCAGTTGCTGGATCTATGATAAC	31825
Qy <td>223</td> <td>GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleSerValIys</td> <td>242</td>	223	GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleSerValIys	242
Db	31826	GAGGGAATTGCCATCTTTGCTCTTCAATTCACGATATTATTTATGGTGAATCGCTCAAG	31885
Qy <td>243</td> <td>ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAla</td> <td>262</td>	243	ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAla	262
Db	31886	ACCGGATCGGATATATGTCGCTTCGTTATGTCGCCCTTTTCATCTTTTACATGTTTCGCA	31945
Qy <td>263</td> <td>TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeu</td> <td>282</td>	263	TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeu	282
Db	31946	TGGGAGGATATGTTTTCATTAACATTGATTCACCTTCAGCTTCGCTTCATCATC	32005
Qy <td>283</td> <td>Met</td> <td>284</td>	283	Met	284
Db	32006	ATGGGTAATTAATAAAAAAATTAATGCAATTTATCATGTCGCAAAAAATTTTATTTCAGG	32065
Qy <td>285</td> <td>-ArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLe</td> <td>304</td>	285	-ArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLe	304
Db	32066	TCGCTACTCGTCCCGTCTGTTGTTTCTTACACATCGTCTTATTGCTTGGCCATCATCCT	32125
Qy <td>304</td> <td>uSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAl</td> <td>324</td>	304	uSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAl	324
Db	32126	CTCGATGCAAGTTTCCATTTGTAGGATTCACCACTTCGTACTTCCGAGCACATGCCAGC	32185
Qy <td>324</td> <td>aAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLe</td> <td>344</td>	324	aAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLe	344
Db	32186	TTTGGAGTATTGGTCTTCTTCAAATTTGCTCTCATGCATGCTAGCTAGCTCGCAACCGCAT	32245

```

Db      1262 TTTCGCACTCCCGAGCTTTGACCGAACACGTAATGCTGAGATTGGAATAAAGACATT 1321
QY      752 LysPheLysHisLeuGluGluAlaPheThrSerGluHisTriPheValArgIleTyrLys 771
Db      1322 AAATTCAGCATTTTGGAGGAAGCTTTTACATCAGACAGCCTGGCTTGCAGGATATATAA 1381
QY      772 VallysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAsnIlePhe 791
Db      1382 GTGAAGCACCCTGACACAGGAGACACTAGGTCAACAACTCGAGTCACCAACATCGTC 1441
QY      792 ProLysGlnLysTyrLeuSerLysLysThrLysArgLysArgGlyTyrIleLysAsn 811
Db      1442 CCCAACACAGAGATTTTGTCAAGAGACACTACTAAAGGAAGGCTGCTACGTAAAAAT 1501
QY      812 LysLeuValPheLysLysGlyLysLysLysLysLysLysLysLysLysLysLysLys 826
Db      1502 AAGCTAGTGTTTAAAGAAAGGCAAGAGACCTCTAAGAGAGACTGTT 1546
RESULT 14
LOCUS   U13019 54118 bp DNA linear INV 13-JAN-2003
DEFINITION Caenorhabditis elegans cosmid T12A2, complete sequence.
ACCESSION U13019
VERSION   U13019.1 GI:1912482
KEYWORDS HTG.
SOURCE   Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
          Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
          Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 54118)
AUTHORS   Waterston,R.
TITLE     Genome sequence of the nematode C. elegans: a platform for
JOURNAL   investigating biology. The C. elegans Sequencing Consortium
MEDLINE   Science 282 (5396), 2012-2018 (1998)
PUBMED   99069613
REFERENCE 2 (bases 1 to 54118)
AUTHORS   Latreille,P.
TITLE     The sequence of C. elegans cosmid T12A2
JOURNAL   Unpublished (2001)
REFERENCE 3 (bases 1 to 54118)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (06-MAR-1996) Genome Sequencing Center, Washington
          University
REFERENCE 4 (bases 1 to 54118)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (26-JUN-1998) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 54118)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (03-AUG-2001) Department of Genetics, Washington
          University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
          Louis, MO 63110, USA
REFERENCE 6 (bases 1 to 54118)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (19-APR-2002) Department of Genetics, Washington
          University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
          Louis, MO 63110, USA
REFERENCE 7 (bases 1 to 54118)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (28-AUG-2002) Department of Genetics, Washington
          University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
          Louis, MO 63110, USA
REFERENCE 8 (bases 1 to 54118)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (21-NOV-2002) Department of Genetics, Washington

```

University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
 9 (bases 1 to 54118)
 Waterston,R.
 Direct Submission
 Submitted (13-JAN-2003) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
 On Jul 27, 1995 this sequence version replaced gi:529354.
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1HQ, England
 email: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=T12A2;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C23G10, 200 bp overlap; the 3' cosmid is C18F10, 15000 bp overlap. Actual start of this cosmid is at base position 1 of T12A2; actual end is at 39122 of T12A2.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFome cloning project (<http://wormdb.dici.harvard.edu/>), similarity to other proteins from BlastX analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tSNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES
 source
 1..54118
 /organism="Caenorhabditis elegans"
 /mol_type="genomic DNA"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="III"
 /clone="T12A2"
 3241..6926
 /gene="hot-4"
 join(3241..3365,3586..3749,6392..6561,6795..6926)
 /gene="hot-4"
 /standard_name="T12A2.16b"
 /note="coded for by the following C. elegans cDNAs:
 yk1067e11.3, yk1325d06.3"
 /codon_start=1
 /product="Homolog of odr-2 (two) protein 4, isoform b"
 /protein_id="AA15601.1"
 /db_xref="GI:20198879"

gene

CDS

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenesScan gene prediction.

FEATURES

source
1..1734
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech II"
/db_xref="taxon:10090"
/clone="IMAGE:358725"
/tissue_type="mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MNV."
/clone_lib="NCI CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..1734
/gene="1300006C19Rik"
/note="synonym: 3imp"
/db_xref="LocusID:68292"
/db_xref="WGI:1915542"
1..1549
/gene="1300006C19Rik"
/codon_start=2
/product="1300006C19Rik protein"
/protein_id="AAH03206.1"
/db_xref="GI:13096824"
/db_xref="LocusID:68292"
/db_xref="WGI:1915542"
/translation="GFQPIRTSEHMAAGVALLQAVFLVLDLTKQEPQLFELGVSLAAGAVFLSVLYLVYGTGAPWSGRFYSIMDTGYAKIHPIIASVSEHQPTTWSFPFDHLILVTFPAGLNFCKINDEKRVFVALYSAISAVTFAGVWVRLMLILFVVCMLSAIATNSVFELGDDKRNPPVEDSDSDKRNPNLYDAGKVRKHVTEQEKPEEGLPNKISIVTLMMLLMFAVCHTWTSNAYSSPVSLASYNHDTNILDPREAYFLWQNTDEHARVMSWMDYQIAGMANRTLLVNDNWNHSHIALVKGAMSSNETAAVKIMSDVDLVLFVFGVYGSDDINKFLMWRIAEGHEPKDIREGYFTQQSEFRVDKAGSPDLLMLVKKSYRFGEMQDLDFRTPPGDFORTNAETGNKDIKFKELEAFTSEHLVRLYVKVAPDNETLGHKPRVNIIVPKQYLSKTKTKRQGVNKNLVFKKGKTSKKTIV"

CDS

misc_feature
1..1249
/gene="1300006C19Rik"
/note="STT3; Region: Oligosaccharyl transferase STT3 subunit. This family consists of the oligosaccharyl transferase STT3 subunit and related proteins. The STT3 subunit is part of the oligosaccharyl transferase (OTase) complex of proteins and is required for its activity. OTase transfers a lipid-linked core-oligosaccharide to selected asparagine residues in the ER"
/db_xref="CDD:pfam02516"

ORIGIN

Alignment Scores:

Pred. No.: 1.03e-205 Length: 1734
Score: 2678.00 Matches: 505
Percent Similarity: 98.45% Conservative: 2
Best Local Similarity: 98.05% Mismatches: 8
Query Match: 60.85% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x BC003206 (1-1734)

Qy 312 GlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuLeu 331
Db 2 GGATTTACGCCAATCAGAACAGCGAGGCACATGCGAGTGTCTTTGGCTGCTG 61
Qy 332 GlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrIysGlnGluPheGlnThr 351
Db 62 CAAGCTTACGCTTTTTCAGTATCTGAGACCGGTTGACAAACAGGAGTTCCAGACC 121
Qy 352 LeuPheLeuGlyValSerLeuAlaAlaGlyValPheLeuSerValIleTyrLeu 371
Db 122 CTTTCTTTTGGGTGTCTACTAGTCAGGCGCTGTGTCTTCTTAGTGTCATCTATCG 181

Qy 372 ThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTyrPheThrGly 391
Db 182 ACATACACAGGTATATATTCACCATGGAGTGGCAGGTTTATTACATATGGGATCTGGG 241
Qy 392 TyrAlaIleValIleHisIleProIleIleAlaSerValSerGluHisGlnProThrTrp 411
Db 242 TATGCAAAATACACATTCCTCAATATTGTCATCAGTGTCTGAACATCAGCCTACGACATGG 301
Qy 412 ValSerPhePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPhe 431
Db 302 GTGCTTTCTCTTTCATCATATCTTCTGTATCTACCTTCCCGCAGGCGCTATGGTTC 361
Qy 432 CysIleIysAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaVal 451
Db 362 TGCATCAAAAATATCAACGATGAAGAGATTATTTGCGTCTGTATGCGATCAGTGCTGTC 421
Qy 452 TyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeu 471
Db 422 TACTTTGCGGAGTGTATGGTGGCTGTATGCTGACTCTGACCCCGGTCGTCTGCATGCTG 481
Qy 472 SerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGlyAspMetLeuArgGlu 491
Db 482 TCGGCCATCGCTTCTCCATGTTTTTGGAGCACTATTTGGGGGATGACATGAAGAAGGAA 541
Qy 492 AsnProProValGluAspSerSerAspGluAspLysArgAsnGlnGlyAsnLeuTyr 511
Db 542 AACCCACCTGTGGAGGACAGCAGTGTATGAGGATGACAAAAGAAACCCAGGAACTTGTAT 601
Qy 512 AspIysAlaGlyIysValArgIysHisAlaThrGluGlnGluIysThrGluGluGlyLeu 531
Db 602 GCAAGGCGAGTAAAGTGGAGAGCATGTGACAGAGCAGAGAGAAACCTGAAGAGGGTTG 661
Qy 532 GlyProAsnIleIysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAla 551
Db 662 GGCCCCAAATCAAAAGCATTTGACCATGCTGTATGCTCATGCTCTGATGATGTTCCGG 721
Qy 552 ValHisCysThrTrpValThrSerAsnAlaTyrSerProSerValValLeuAlaSer 571
Db 722 GTCCACTGCACTGGGTGCACAGCAACGCTACTCCACTCCAGTGTGGTCTTGGCTCC 781
Qy 572 TyrAsnHisAspGlyThrArgAsnIleLeuAspPheArgGluAlaTyrPheTrpLeu 591
Db 782 TAGCAATCATGATGTTACCGAGTAATATATATGATGATATTTAGAGAGCGGTATTTGGCTG 841
Qy 592 ArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIle 611
Db 842 AGACAAAACACGGATGAACACCGCCGGGTCTATGCTGTGGTGGACTACGGCTATCAGATT 901
Qy 612 AlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 631
Db 902 GCTGGCATGGCCCAACAGGACCACTCTGGTGGATAACAACACCTGGAAACAACAGGCACATC 961
Qy 632 AlaLeuValGlyIysAlaMetSerSerAsnGluThrAlaAlaTyrIleMetArgThr 651
Db 962 GCATGGTGGAAAAAGCTATGCTTCCATGAACCGCGCTATATAAATCATGAGGTCC 1021
Qy 652 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAsp 671
Db 1022 CTTGATGTGATATGTTGTTGTTATTTTCGGAGAGGTGATTTGGCTATTTCCGGGAGCAT 1081
Qy 672 IleAsnIysPheLeuTrpMetValArgIleAlaGluGlyGluHisProIysAspIleArg 691
Db 1082 ATCAACAAGTTCTCTGATGTTGTCAGGATAGCTGAAGGGGAGCATCCCAAGACATCCGG 1141
Qy 692 GluSerAspTyrPheThrProGlnGlyGluPheArgValAspIysAlaGlySerProThr 711
Db 1142 GAAGGTGACTATTTTACCAGCAGGAGAGATTCCGAGTAGACAAGCTGGGTCTTCTTACT 1201
Qy 712 LeuLeuAsnCysLeuMetTyrIysMetSertyrTyrArgPheGlyGluMetGlnLeuAsp 731
Db 1202 CTGTAAACTGCCTTATGATAAAATGTCATACTACAGATTGGAGAAATCGACCTAGAT 1261
Qy 732 PheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnIysAspIle 751

```

Qy 493 ProProValGluAspSerSerAspGluAspAspLysArgLysGlnGlnIleuLeu----- 510
Db 20141 GATGAAGCTGAGGATTCATTGA- GAAGAAGACGCTGTACGACAAAGGTGAGTTCTTACTA 20083
Qy 511 -----TyrAspLysAlaGlyLysValArg 518
Db 20082 ACACATCATCCATGGATTGATTTTAAATAATCATCGCATTTGCAGGCTGCAGCATG--- 20026
Qy 519 LysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsnIleLysSer 537
Db 20025 AGCATCTGACTAAAGCATGATGACCCAGCAGGATGACTGCGCTGACGCTCCAACTCGAAGATG 19966
Qy 538 IleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
Db 19965 ATTGTTATTGCGCGTCTTAATGCTGTGTGATGATGTTGCTGTGCCACTGCACGTGGGTG 19906
Qy 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
Db 19905 ACAGCAATGCTACTCCAGTCCCTCCATTGCTGTGGCTTTCACACACAGTCAAGATGG 19846
Qy 577 -----Thr 577
Db 19845 TAAGCACCAAGTAGCATCATGCTCTTTTAAACCTATCTAACTTTTCTCCACTAGATCC 19786
Qy 578 ArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGlu 597
Db 19785 CGCAACATTTAGACGATTTAGAGAGCTTACTACTGCTTTCGAGAACACTGCGCAT 19726
Qy 598 HisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 617
Db 19725 GATGTCGCGTATGCTCTGTGGTATTCGATACCATACAGATAGCGGAATGGCAACAGA 19666
Qy 618 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 637
Db 19665 ACAGCGCTAGTGATTAATACGTTGAGACCAATAGTCATAGCTGTGTTGCGAAGCA 19606
Qy 638 MetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrVal 657
Db 19605 ATGCTTCAACCGAGAGAGAGTCTCAACAAATATGACATCTCTTGCAGTGGACTACGTT 19546
Qy 658 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 677
Db 19545 TTGGTGATCTTGGCGGTGATCGGTATTCGCGCATGATATCAACAGTCTCTGTGG 19486
Qy 678 MetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThr 697
Db 19485 ATGGTCGGAATTCGTCGAGGAGAGCATCCCAAGGACATTAAGGAAGCGGATTAATC 19426
Qy 698 ProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMet 717
Db 19425 GACCGGCTGATTCAGGGTAGATCCCGAGGTGCTCCGCGCTCTCACTGCCTTATG 19366
Qy 718 TyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThr----- 734
Db 19365 TACAATTAAGCTACTACAGATTCGGGAATTTGAAGTTGGACTACAGGTA-AGCGGAAC 19307
Qy 735 -----ProGlyPheAsp 739
Db 19306 ATTCTCTCAGGTACGATGCACACTAATGCTGTTTTCAGAGGTCCATCTGATATGAT 19247
Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysIleGluGluAla 759
Db 19246 CGCACACGTAAGCGCGTCAATGGGAATTAAGCACTTCGATCTGACCTCGAGGAGGCGC 19187
Qy 760 PheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAsp-----Asn 777
Db 19186 TACACCACAGAACACTGCTGCTTGTTCGATCTATAGGCTGAAGAAGCGCATGAGTCAAT 19127
Qy 778 ArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeu 797
Db 19126 AGACCATCACTGAAGACCAAGGAGAGACG-----ATTCTCCAGCAACTTCATT 19076

```

798 SerLysLysThrThr----- 802
 19075 TCAGAGAAAGGT-AGTGTACCTCCCTGCAGCTCCCAATGAACCTCATTAAACAATCTGATTTT 19017
 803 -----LysArgLysArgGlyTyrIleLysAsnLysLysLeuValPheLysLysGlyLys 819
 19016 GCAGAACCTAAGGCTCGCAAGGCTTACATAGAAACCGCCGTTGTGTAGGGAATA 18957
 820 LysIleSerLysLysThr 825
 18956 CGAACCTTGAATAAACC 18939
 RESULT 13
 BC003206 1734 bp mRNA linear ROD 12-NOV-2003
 LOCUS Mus musculus RIKEN CDNA 1300006C19 gene, mRNA (cDNA clone
 DEFINITION IMAGE:3587254), partial cds.
 BC003206
 VERSION BC003206.1 GI:13096823
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1734)
 AUTHORS Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L.,
 Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S.,
 Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J.,
 Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J.,
 McKernan K.J., Malek J.A., Gunaratne P.H., Richards S.,
 Worley K.C., Hale S., Garcia A.M., Gay D.J., Hulyk S.W.,
 Villalón D.C., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Hellon E., Kettelman M., Madan A., Rodriguez S.,
 Sanchez A., Whitting M., Madan A., Young A.C., Shvachenko Y.,
 Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D.,
 Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 Butterfield Y.S., Krzywinski M.I., Skalska U., Smaluk D.E.,
 Schnerch A., Schein J.E., Jones S.J. and Varra M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 REFERENCE 2 (bases 1 to 1734)
 AUTHORS Strausberg R.
 TITLE Direct Submission
 JOURNAL Submitted (20-FEB-2001) National Institutes of Health, Mammalian
 MEDLINE Gene Collection (GNC), Cancer Genomics Office, National Cancer
 PUBMED Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: angbcm@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounsged, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 10 Row: 0 Column: 16


```

Db 128698 TACACAGACACTGCTGTTGCGCACTATAGGTGAAGACCGCATGATGATCAAT 128699
Qy 778 ArgGluThrLeuAspHisLeuProArgValThrAsnLeuPheProLysGlnLysTyrLeu 797
Db 128638 AGACCATCACTGAAGACCAAGAGAGAACG-----ATTCTCCAGCAAACTTCAT 128588
Qy 798 SerLysLys-----800
Db 128587 TCGAGAAAGTAGTACCTCCAGCTCCCAATGCAATCTATTAAACATCTGATTTTG 128528
Qy 801 ----ThrThrLysArgLysArgGlyTyrLeuLysAsnLysLeuValPheLysLysGlyLys 819
Db 128527 CAGAACTCTAACCGCTCGAAGCGGTACATACGAAACCGCGGTGTGTGTAAGAA 128468
Qy 820 LysLysSerLysLysThr 825
Db 128467 CGAACCTTGAAATAAC 128450

RESULT 12
AC018145/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC018145.1 75650 bp DNA linear HTG 09-DEC-1999
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
AC018145
GI:6553046
HTG: HTGS PHASE2.
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 75650)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10213542 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..75650
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Alignment Scores:
Pred. No.: 1,61e-212 Length: 75650
Score: 2791.50 Matches: 556
Percent Similarity: 76.12% Conservative: 88
Best Local Similarity: 65.72% Mismatches: 115
Query Match: 63.43% Indels: 91
DB: 2 Gaps: 10

US-10-028-384-2 (1-826) x AC018145 (1-75650)
Qy 63 SerGlnProAlaGlyTyrGlnSerLeuLeuSerPheThrLeuPheLeuAlaTyrLeu 82
Db 21460 AGCAAGTGGCTGGCTACAGACGCTATACCTTCGCCATCTCTGTAATCGCTGGCTG 21401
Qy 83 AlaGlyPheSerSerArgLeuPheAlaValLeuArgPheGluSerLeuLeuHisGluPhe 102
Db 21400 GCGGATTTTCCTCTCGCCCTTCGCCGCTATCGCTTCGAGTCGATATCCATGAGTTT 21341
Qy 103 AspProTyrPheAsnTyrArgSerThrHisLeuAlaSerHisGlyPheTyrGluPhe 122
Db 21340 GATCCGTGGTCACTACCGGCCACCGCTACATGGTGGAGATGTTGGTACAACTTC 21281
Qy 123 LeuAsnTyrPheAspGluArgLalaTyrTyrProLeuGlyArgLleValGlyThrVal 142

```


Shu, S.O., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W.M.,
Rubin, G.M. and Lewis, S.E. *Drosophila melanogaster* euchromatic genome: a
systematic review
Genome Biol. 3 (12), RESEARCH0083 (2002)
22426069
12537572
4 (bases 1 to 227219)
Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J.,
Svirkas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E.,
Rubin, G.M., Ashburner, M. and Celniker, S.E.
The transposable elements of the *Drosophila melanogaster*
euchromatin: a genomics perspective
Genome Biol. 3 (12), RESEARCH0084 (2002)
22426070
12537573
5 (bases 1 to 227219)
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
6 (bases 1 to 227219)

FlyBase
Direct Submission
Submitted (06-SEP-2002) University of California Berkeley, 539 Life
Sciences Addition, Berkeley, CA 94720, USA
7 (bases 1 to 227219)

FlyBase
Direct Submission
Submitted (10-MAR-2004) FlyBase, Harvard University, Biological
Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA
On or before Sep 18, 2002 this sequence version replaced
gi:10729289, gi:7301241

FEATURES

source

1..227219
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="3R"
complement(149..6378)
/gene="ssh"
/locus_tag="CG6238"
/map="96810-96811"
/db_xref="FLYBASE:FBgn0029157"
complement(join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..6378))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RA"
/db_xref="FLYBASE:FBgn0029157"
complement(join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RA"

mRNA

complement(join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))

CDS

1..396-215
Pred. No.: 2838.00
Score: 76.24
Percent Similarity: 65.84
Best Local Similarity: 64.49
Query Match: 3
DB: 10
US-10-028-384-2 (1-826) x AB003750 (1-227219)

Qy 63 SerGlnProAlaGlyTrpGlnSerLeuSerPheThrIleuPheLeuAlaTrpLeu 82
Db 130973 AGAAGTGGCTGGCTACAGACCCATACCTCCGCATCTCTGATTCGCTGGGTG 130914
Qy 83 AlaGlyPheSerSerArgLeuPheAlaValIleAtrgPheGluSerIleIleHisGluPhe 102
Db 130913 GCCGATTTTCCCTCGCTCTCTCGCGCTTCGCGTTCGCTGATTCATCATGAGTTT 130854

KATMMSVDLDEVTSKYIRLEBEILDMDLGEVKSFIIDAEMLVILGQMDAPTKIRHHV
LGSEWNASLLELQKNGVPHILNVREINDNFPFGTFEYFNVRVVDDKTNLLKXWDDI
FRITTKAAGSKVLCHCKGVSRSASVVIATAMKAYQWERQQALDEHVKKRRSCIKPN
KFNLEQLEYSGLDAMKNEKLQSKSETNLKADYKARLLFGSEPTPLIQAQKAS
KSTGEAGVTPDGEEDGSMRRSIAQSKQRMVRSSTSPKTATVAVTQQQSQSME
NLTPERSVAEPEKNRPFPGSNGENTSVTQNLVHIQKHTPLSVTRIHLEARADQL
POOPVMTSLTKLITOTSHLGKSVSGSGNIDRRDSCSVFSDQSVFPAKDEGEK
RORRKHSTESLPGSGIVLPTQQQKQSNAILRPRGTQRQELPRHASWGSQDN
RCCLPORTSSGSYDSNRNTTAIFSGVLDLKRSSNCNVIEGVAVAVPEIYVGVGTV
KCKOKLESTSLKKECOSEQLLEAVDAGORCPSLYBSASASPRORPLRCN
SEELMTSIEKNKTPPDHHEATVLRVQGTLDQRIQSNVQLKKNFPAKAGVG
TGGGGGGAAGSGSSSTIATSVAVAGCSANPKGKTTSHQSLEPSSFAVAHANVSA
ASNSASNSDSVSENCNRNKLGHVHYQITTTSCSSQAATPTACALPAGTA
SNHSCTTTTTFKQRLSYDRSLNRSHTDYSRPPPPVPRVVSNSVQHQOQ
QQOQLQPEHQOQVIRLAHQKTHPLSRAPKTFNTAAYNTM"
complement(join(875..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5095))
/gene="ssh"
/locus_tag="CG6238"
/codon_start=1
/product="CG6238-PB"
/protein_id="AAN14027.1"
/db_xref="GI:23172231"
/db_xref="FLYBASE:FBgn0029157"

CDS

1..227219
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="3R"
complement(149..6378)
/gene="ssh"
/locus_tag="CG6238"
/map="96810-96811"
/db_xref="FLYBASE:FBgn0029157"
complement(join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..6378))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RA"
/db_xref="FLYBASE:FBgn0029157"
complement(join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RA"

gene

complement(join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))

mRNA

complement(join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))

mRNA

complement(join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))

Alignment Scores:

Pred. No.: 1.396-215 Length: 227219
Score: 2838.00 Matches: 557
Percent Similarity: 76.24% Conservative: 88
Best Local Similarity: 65.84% Mismatches: 114
Query Match: 64.49% Indels: 89
DB: 3 Gaps: 10

US-10-028-384-2 (1-826) x AB003750 (1-227219)

Qy 63 SerGlnProAlaGlyTrpGlnSerLeuSerPheThrIleuPheLeuAlaTrpLeu 82
Db 130973 AGAAGTGGCTGGCTACAGACCCATACCTCCGCATCTCTGATTCGCTGGGTG 130914
Qy 83 AlaGlyPheSerSerArgLeuPheAlaValIleAtrgPheGluSerIleIleHisGluPhe 102
Db 130913 GCCGATTTTCCCTCGCTCTCTCGCGTTCGCGTTCGCTGATTCATCATGAGTTT 130854

Brenhoff, C., Champagne, M., Chavez, C., Chew, M., Ciesiolka, L., Doolittle, A., Fardoun, E., Gallo, P., Gess, R., A. H. Kitzner, N.L., Hokits, P.A., Houston, K.A., Hummels, S., Karas, K., Keefe, J., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Mohrati, A.B., Mosher, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poorn, L., Sequeira, A., Sethi, H., Snir, E., Svirekas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

kwu44,9.0.0.
Direct Submission
Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 24, 2001 this sequence version replaced qi:7208934.

Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdpc@fruitfly.edu.

```

to bug@ruaffcity.berkeley.edu.
Location/Qualifiers
1. .181132
   /organism="Drosophila melanogaster"
   /mol_type="genomic DNA"
   /strain="y; cn bw 8D"
   /db_xref="taxon.7227"
   /chromosome="3R"
   /map="96B-96B"
   /clone="BACR03115 (D765)"
   /clone_lib="RPCC-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ESORI in
pBACe3.6)"

```

ORIGIN

Alignment Scores:		
Pred. NO.:	1.01e-215	Length:
Score:	2838.00	Matches:
Percent Similarity:	76.24%	Conservative:
Best Local Similarity:	65.84%	Mismatches:
Query Match:	64.49%	Indels:
DB:	3	Gaps:
US-10-028-384-2 (1-826)	x	AC008206 (1-181132)

QY 798 SerLysLys----- 800
 Db 1430 TCAGAAAGGTAGTCTACCTCGAGCTCCGATGACCTCATTAACATGATGTTTC 1371

QY 801 ----ThrThrLysArgLysGlyTyrLysLysLysLysLysValPheLysLysGlyLys 819
 Db 1370 CAGAACTCTAAGCGTCGAGCGGTACATACGAAACCGCGGTGTGTGTAAGGAAAA 1311

QY 820 LysLysSerLysLysThr 825
 Db 1310 CGAACTTCGAATAAACC 1293

RESULT 9

AC007853/c

LOCUS 162921 bp DNA linear INV 27-FEB-2001
 DEFINITION Drosophila melanogaster, chromosome 3R, region 96B-96C, BAC clone
 BACR03L02, complete sequence.

ACCESSION AC007853
 VERSION 4 GI:13129410
 KEYWORDS HTG.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 162921)

AUTHORS Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
 Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
 Rogers, J., An, H., Baldwin, D., Banazon, J., Beeson, K.Y., Busam, D.A.,
 Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,
 Dodson, K., Dorsett, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D.,
 Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
 Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
 Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
 McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunco, J.,
 Pacleeb, J., Paragov, V., Park, S., Patel, S., Pfeiffer, B.,
 Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
 Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
 Zaveri, J.S., Smith, H.O., Rubin, G.M., and Venter, J.C.
 Sequencing of Drosophila chromosome 3R, region 96B-96C
 Unpublished

TITLE

JOURNAL

REFERENCE

AUTHORS 2 (bases 1 to 162921)
 Celniker, S.E., Agbavani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Burenhoff, C., Champ, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.D.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleeb, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.

TITLE Direct Submission
 JOURNAL Submitted (17-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT On Feb 27, 2001 this sequence version replaced gi:5670481.

Sequence submitted by:
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdgs@fruitfly.berkeley.edu.

FEATURES

source

1. 162921
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /map="96B-96C"

/clone="BACR03L02 (D766)"
 /clone_lib="RPCI-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial ECORI in
 pBACe3.6)"

ORIGIN

Alignment Scores:
 Pred. No.: 8,67e-216 Length: 162921
 Score: 2838.00 Matches: 557
 Percent Similarity: 76.24% Conservative: 88
 Best Local Similarity: 65.84% Mismatches: 114
 Query Match: 64.49% Indels: 89
 DB: 3 Gaps: 10

US-10-028-384-2 (1-826) x AC007853 (1-162921)
 QY 63 SerGlnProAlaGlyTrpGlnSerLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
 Db 108454 AGCAAGGTGGTGGTGTACAGCAGCCATATCAGCTTCGCCATCTCTAATCCCTGGCTG 108395
 QY 83 AlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
 Db 108394 GCCGATTTTCCTCGCTCTTCGCGGTATCCGCTTCAGTCGATATCATCATGAGTTT 108335
 QY 103 AspProTrpPheAsnTrpArgSerThrHisHisLeuAlaSerHisGlyPheTrpGluPhe 122
 Db 108334 GATCCGTTGGTTCACCTACCGGCGCACCGCTACATGGTCCAGATGGTTGGTACAACTTC 108275
 QY 123 LeuAsnTrpPheAspGluAlaTrpTrpProLeuGlyArgIleValGlyThrVal 142
 Db 108274 CTCAACTGGTTCAGCAGGCGCATGGTATCCGCTCGGAGGATTTGGGGGTACCGTC 108215
 QY 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle 162
 Db 108214 TATCCCGGCTGTATGATTACGTCGCGGGAATCCATGGCTGCTGCATCTCAACATA 108155
 QY 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 Db 108154 CCGGTCCATATTCGTGACATCTTCGCTGCGGTTCCTGGGCGCATCTTCAGTGGCTGACCTCC 108095
 QY 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAlaGlyLeuAlaAla 202
 Db 108094 ATCTCCACTACCTGTGTGACCAAGAGAGCTGTGGTTCGGGGGCGCGGCTCTTCGCCGCC 108035
 QY 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAsp 222
 Db 108034 AGCTTCATCGCCATCGTGTGCTGTGTACATCAGTAGTGGTGGTGGTACGATACGATAAC 107975
 QY 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpValValSerValLys 242
 Db 107974 GAGGCAATTGCCATATTTCGCCCTTCAGTTCACCTACTTCTGTGGGTGGCTCAGTGAG 107915
 QY 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMetValSerAla 262
 Db 107914 ACTGGATCCGTGTGTGTGTCGGCGCGCAGCGCGGTGTGTCTACTTCTACATGTGTGCGCC 107855
 QY 263 TrpGlyGlyTrpValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeu 282
 Db 107854 TGGGTGGCTACGTGTGTATCATCACTATACCTATACCTTCGACGCTTCTGTACTGCTCAT 107795
 QY 283 MetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTrpIleValGlyLeu 302
 Db 107794 ATGGCAGGTACTTCGCGCGGTCTGTGTGACCACTACAGCAGCTTCTACATCTCGGAGCTG 107735
 QY 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 Db 107734 CTGTTCTCCATGAGATCCCTTCGTGGGATTCACCGATACGACCGACGATGACACATG 107675
 QY 323 AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
 Db 107674 GCTGCGCTGGGAGTGTGTGTGCTCTTATGCGCGGTGCGCACCGCTTGGCCATTTGCGATCC 107615
 QY 343 ArgLeuThrLysGlnGlnPheGlnThrLeuPhePheLeuValSerLeuAlaAlaGly 362

Db 3636 CTCAACTGTTTCGACGCGCGCATGGTATCCGTCGGCAGGATTCGGCGCGTACCGGTC 3577
 Qy 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle 162
 Db 3576 TATCCCGGCTGATGATACGTCGCGCGGATCCATTCGTCGTGCACGTACTCAACATA 3517
 Qy 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 Db 3516 CGGTCATATTCGTCATCTCGTCTCTCGCGCGCATCTTCAGTCGCTGACCTCC 3457
 Qy 183 IleSerThrPheLeuLeuThrArgGlyLeuIleTrpAsnGlnGlyLeuLeuAla 202
 Db 3456 ATCTCCACCTACCTGCTGACCAAGAGCTGTGGTCGGCGCGCGCTCTTCGCGCGC 3397
 Qy 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspIle 222
 Db 3396 AGCTTCATCGGCATCGTTCGCTGGGTATCATCAGTAGTCGGTGGCTGGATCGTACGATAAC 3337
 Qy 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleValSerValIys 242
 Db 3336 GAGGCGATTCGATATTCGCTCGCATTCACCTACTCTCTGTTGGTGGCTCAGTGAAG 3277
 Qy 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
 Db 3276 ACTGGATCCGCTGTTCTGGTCGGCGCGACCGCTTGTCTACTTCTACATGGTGTCCGCC 3217
 Qy 263 TrpGlyGlyTyrValPheIleAsnLeuIleProLeuHisValPheValLeuLeuLeu 282
 Db 3216 TGGGGTGGCTACGTGTTTCATCATCACTGATACCCCTGTCACCTCTCTGCTACTGCTCAT 3157
 Qy 283 MetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
 Db 3156 ATGGCAGGTACTCGCGCGCTCTCTGCTGCTGACGAGTACAGCACTTCACATCTCTGGAGCTG 3097
 Qy 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 Db 3096 CTGTTCTCATGCGATCCCTCTGTCGGATTCACACCGATACGACCATGACACATG 3037
 Qy 323 AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
 Db 3036 GCTGCGCTGGAGTGTTCGCTCTATGCGCGTGGCCACCTTCGCGCATTTGCGAGTCC 2977
 Qy 343 ArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuValSerLeuAlaAlaGly 362
 Db 2976 GTGCTGTGCGGACGAGTTCGGAAGCTGTTCATGTCGCGCGATTCGTCGGCGCT 2917
 Qy 363 AlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProThrPheGly 382
 Db 2916 GCGCTGTTTGTGGCGCTGCTGCTCACCATGCTGGCGCTGTGTGGCGCGCTGGA 2857
 Qy 383 ArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSer 402
 Db 2856 CGCTTCATCTGCTGGGATCTGCTACGCGCAGATCCACATTCCTCATTCATGCTATCC 2797
 Qy 403 ValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHisIleLeuVal 422
 Db 2796 GTGCGGAGATGACGCCACCATGCTGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTG 2737
 Qy 423 CysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPhe 442
 Db 2736 TGGCGCTTCCTCCAGTGGGAGTGTGCTACTGCATCAGCAGATCAACGCGCGCTTTC 2677
 Qy 443 ValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 462
 Db 2676 GTGGTGTGTCAGGCGATCAGTGGCGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2617
 Qy 463 ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482
 Db 2616 ACCCTCAGCGCGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2557
 Qy 483 TyrLeuGlyAspAspMetLysArg-----GluAsn 492
 Db 2556 TTCCTGCAAGAGGATTCGCTTAAGCGAATGGCGACAGCCATAGACGCGACCGGAGTGT 2497

Qy 493 ProProValGluAspSerSerAspGluAspAspLysArgAsnGlnGlyAsnLeu----- 510
 Db 2496 GATGAAGCTGAGGATTCATTGA-GAAGAAGACGCTGTACGACAGGTGAGTCTTCTACTA 2438
 Qy 511 -----TyrAspLysAlaGlyLysValArg 518
 Db 2437 ACACACATCCATGGATTTGATTTAAATACATCGCATTCGAGCTGGCAAGCTG--- 2381
 Qy 519 LysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsnIleLysSer 537
 Db 2380 AAGCATCGTACTAAGCATGATGCCAGCAGAGTACTGGCGTCAGCTCCAACTCAAGAGT 2321
 Qy 538 IleValThrMetLeuMetLeuMetLeuMetPheAlaValHisCysThrTrpVal 557
 Db 2320 ATTGTTATTGTCGCTTCTAATGCTGTGTGATGATGTTGCTGCTGCTGCTGCTGCTGCTG 2261
 Qy 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
 Db 2260 ACCAGCAATGCTACTTCCAGTCCCTCCATGCTTGGCTTTCCACACAGTCAAGATGG 2201
 Qy 577 -----Thr 577
 Db 2200 TAGCACCAAGTAGCATCATGCTCTCTCTTAAACCTATCTAACCTTTCTCCACTAGATCC 2141
 Qy 578 ArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGlu 597
 Db 2140 CGCAACATTTAGACGATTCAGAGAGGCTTACTACTGGCTTCGACAGACACTGCCGAT 2081
 Qy 598 HisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 617
 Db 2080 GATGCTCGGCTGATGCTTGGTGGGATTCAGGATACAGATAGCGGATGGCAACAGA 2021
 Qy 618 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 637
 Db 2020 ACCAGCTAGTGGATAATAATACGTGGAAACATAGTACATAGCGCTGTTGGCAAGCA 1961
 Qy 638 MetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrVal 657
 Db 1960 ATGCTTCACCGAGGAGAGTCTACGAAATATGACATCTCTGACGTGAGTACTGTT 1901
 Qy 658 LeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 677
 Db 1900 TTGGTATCTTGGCGGTGTGATCGCTATTCGCGATGATATCAACAGTCTCTGTGG 1841
 Qy 678 MetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThr 697
 Db 1840 ATGCTCGAATGCTGAGGAGAGCATCCCAAGGACATTAGGAAGGAGGATTTACTTACC 1781
 Qy 698 ProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMet 717
 Db 1780 GACCGCGGTGAATTCAGGCTAGATCCGAAGGTGCTCCGCGCTGCTCAACTGCTTATG 1721
 Qy 718 TyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThr----- 734
 Db 1720 TACAAATTAAGCTACTACAGATTCGGGAATTTGAAGTTGGACTACAGTA-AGCGGAAC 1662
 Qy 735 -----ProProGlyPheAsp 739
 Db 1661 ATTTCTCTCAGGTAGCGATGCACATACTGCTCTTTTCAGAGGTCTCCATCTGGATATGAT 1602
 Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAla 759
 Db 1601 CGCACACGTAACCGCTCATTTGGGAATAGGACTTCGATCTGACCTACCTCGAGGAGGCC 1542
 Qy 760 PheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAsp-----Asn 777
 Db 1541 TACACACAGAACACTGCTTGTTCGCTATATAGGTCGATAGGTCGATCTGACCTACCTCGAGGAGGCC 1482
 Qy 778 ArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeu 797
 Db 1481 AGACCACTACTGACAGCAAGGAGAGACG-----ATTCCTCCAGCAAACTTCATT 1431


```

Qy 577 ThrArgAenIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAsp 596
Db 1655 TCCGCGACATTTTAGACGATTTTCAGAGAGGCTTACTACTGGCTTTTCGAGAACACCTGCC 1714

Qy 597 GluHisAlaArgValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAen 616
Db 1715 GATGATGCTCCGGTATGCTTGTGGATACCGATACCGATACCGGGAATGGCAAC 1774

Qy 617 ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisAlaIleAlaValGlyLys 636
Db 1775 AGAACGAGCTAGTGGATTAATAACCTGGAAACAATAGTACATACGCTGGTGGCAAG 1834

Qy 637 AlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyr 656
Db 1835 GCATGCTTCAACCGAGAGAGACTCTCAGAAATATGACATCTTTCAGCTGGACATAC 1894

Qy 657 ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 676
Db 1895 GTTTTGGTGATCTTTGGCGGTGTCATCGCTATTCTTGGCGATGATATCAACAAGTTCTTG 1954

Qy 677 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPhe 696
Db 1955 TGGATGTCGGAATTCGTGGAGAGACATCCCAAGGACATTAAGGAAGCGATTACTTT 2014

Qy 697 ThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeu 716
Db 2015 ACCGACCGCGGTGAATTCAGGGTAGATGCCGAAGGTGCTCCGCGCTGTCTCAACTGCCCT 2074

Qy 717 MetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProPro 736
Db 2075 ATGTACAAATTAAGCTACTACAGATTCCGGGAATTGAAGTTGACTACAGAGTCCATCT 2134

Qy 737 GlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeu 756
Db 2135 GGATATGATCGCACACGCTAAGCCGCTCATTTGGGAATGAAGACTTCGATCTGACCTACCTG 2194

Qy 757 GluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAsp 776
Db 2195 GAGGAGGCTACACACAGACATCTGCTGTTCGATCTATAGGTTGAGAGCGGCAT 2254

Qy 777 -----AsnArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGln 794
Db 2255 GAGTTCAATAGACCATCACTCAAGACCAAGCAGAGAACG-----ATTCTCCAGCA 2305

Qy 795 LysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuVal 814
Db 2306 AACTTCATTTGAGAAAGAACTCTAAGGTCGCAAGGGCTACATAGAAACCGCGGTT 2365

Qy 815 PheLysLysGlyLysLysLysSerLysLysThr 825
Db 2366 GTTGTAAAGGAAACCAACCTTGAAATAAAC 2398

RESULT 7
LOCUS CQ589353
DEFINITION Sequence 17111 from Patent WO0171042.
ACCESSION CQ589353
VERSION CQ589353.1 Gi:41648215
KEYWORDS
SOURCE Drosophila sp.
ORGANISM Drosophila sp.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1
REFERENCE
AUTHORS Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
TITLE Detection kits, such as nucleic acid arrays, for detecting the
expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL Patent: WO 0171042-A 17111 27-SEP-2001;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
1..2699
/organism="Drosophila sp."
source

```

```

/mol_type="unassigned DNA"
/db_xref="taxon:7242"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 2,5e-227 Length: 2699
Score: 2950.00 Matches: 557
Percent Similarity: 83.79% Conservative: 89
Best Local Similarity: 72.24% Mismatches: 113
Query Match: 67.03% Indels: 12
DB: 6 Gaps: 7

```

US-10-028-384-2 (1-826) x CQ589353 (1-2699)

```

Qy 63 SerGluProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
Db 107 AGCAAGTGGCTGGCTACAGCAGCTATATCACTTCGCATCTCTGTAATCGCTGGCTG 166

Qy 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
Db 167 GCGGATTTTCTCTCGCTCTTCGCGCTCATCCGTTTCGAGTCGATTATCATCAGATT 226

Qy 103 AspProTrpPheAsnTyrArgSerThrHisHisIleAlaSerHisGlyPheTyrGluPhe 122
Db 227 GATCCGTGGTTCACTACCGGCGCACCGCTACATGGTGCAGATGGTGTGTACATTC 286

Qy 123 LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyThrVal 142
Db 287 CTCAACTGGTTCGACGAGCGCATGGTATCCGCTCGGCAGAGATTGTGGGCGGTACCGTC 346

Qy 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsnThrLeuAsnIle 162
Db 347 TATCCCGGCTGATGATACGTCCGCGGCAATCCATTCGGCTGTCGACGTACTCAACATA 406

Qy 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
Db 407 CCGGTCCATATTCGTACATCTCGCTGTTCCTTCGGCGCGATCTTCAGTGGCGTCACTCC 466

Qy 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuAlaAla 202
Db 467 ATTCCACCTACTCTGTGACCAAGAGCTGTGGTCCGCGGCGCGGCTCTTCGCGGCC 526

Qy 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn 222
Db 527 AGCTTCATCGCCATCTGCTGGCTACATCAGTAGGTCGGTGGTGGTATCGTACGATAAC 586

Qy 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLys 242
Db 587 GAGGCAATGCCATATTCGCTTCAGTTCACCTACTCTCTGTGGTGGTGGTCTAGTGAG 646

Qy 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
Db 647 ACTGGATCCGTTCTGGTTCGGCGCGCAGCGCTTTTGTCTACTTCTACATGGTCTCGCC 706

Qy 263 TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeu 282
Db 707 TGGGTGGCTACGCTGTTCAATCACTGATACCCCTGCAGCTCTTCGTACTGCTCAT 766

Qy 283 MetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
Db 767 ATGGGAGGTACTCGCGCGGTCTGTGTGACCACTACAGACCTTCTACATCTCGGACTG 826

Qy 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
Db 827 CTGTTCTCCATGCAGATCCCTTCGTGGGATTCGACCGATACGACCGACGATGACACATG 886

Qy 323 AlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
Db 887 GCTGCGGTGGAGTGTGTGTCTCTATGGCGGTGGCGCACCTTCGCGCATTTTCAGTCC 946

Qy 343 ArgLeuThrLysGlnGluPheGlnThrLeuPheLeuGlyValSerLeuAlaAlaGly 362
Db 947 GTGCTGTGGCGCAACGAGTTCGGAAGCTGTTTCATCGTCGGGCGGATTCTGTGGTGGCTT 1006

```

Db 2306 AACCTTCATTCGAGAAAGAACTCTAAGCGTCGCAAGGCGCTACATACGAAACCGACCGGTT 2365
 QY 815 PheLysLysGlyValLeuSerLysThr 825
 Db 2366 GTTGTTAGGGAACGACCTTGAATAAAC 2398

RESULT 6
 AX799088
 LOCUS AX799088 2417 bp DNA linear PAT 08-OCT-2003
 DEFINITION Sequence 7 from Patent WO03054008.
 ACCESSION AX799088
 VERSION AX799088.1 GI:37605061
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
 AUTHORS Perreault, C. and McBride, K.
 TITLE Mammalian SIMP protein, gene sequence and uses thereof in cancer
 therapy
 JOURNAL Patent: WO 03054008-A 7 03-JUL-2003;
 COMPATIGENE Inc. (CA)
 FEATURES
 source
 1..2417
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /note="STT3, Acc# AF132552"

ORIGIN

Alignment Scores:
 Pred. No.: 2,148-227 Length: 2417
 Score: 2950.00 Matches: 557
 Percent Similarity: 83.79% Conservative: 89
 Best Local Similarity: 72.24% Mismatches: 113
 Query Match: 67.03% Indels: 12
 DB: 6 Gaps: 7

US-10-028-384-2 (1-826) X AX799088 (1-2417)

QY 63 SerGlnProAlaGlyTrpGlnSerLeuSerPheThrLeuPheLeuAlaTrpLeu 82
 Db 98 AGCAAGTGGTGGTACGAGCGCTAATACCTTCGCCATCTCTGCTTAATCGCGCTGGCTG 157
 QY 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
 Db 158 GCGGATTTCTCTCGCCTCTTCGCGCTATCCGCTTCGAGTCGATTCATCCATGAGTTT 217
 QY 103 AspProTrpPheAsnTyArgSerThrHisLeuAlaSerHisGlyPheTrpGluPhe 122
 Db 218 GATCCGTGGTTCAACTACCGGCCACCGCTACATGTTGCAAGATGTTGGTACAACTTC 277
 QY 123 LeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGlyThrVal 142
 Db 278 CTCAACTGTTTCGACGAGCGCATGTTATCCGCTCGCGAGGATTTGGCGGCTACCGTC 337
 QY 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpLeuAsnThrLeuAsnIle 162
 Db 338 TATCCCGGCTGATGATACGTCGCGGAAATTCATTTGCTGCTGCGACGACTCAACATA 397
 QY 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 Db 398 CCGGTCCATATTCGTGACATCTGCTGTTCTCTGGCGCGATCTTTCAGTGGCTGACCTCC 457
 QY 183 IleSerThrPheLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaA 202
 Db 458 ATCTCACCTACTGCTGACCAAGGAGCTGTGGTCCCGGGCGCGCGCTCTTCGCGGCC 517
 QY 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAsn 222
 Db 518 AGCTTCATCGCATCGTGGCTGGCTACATCAGTAGTGGTGGTGGTGGTGGTGGTGGTGGT 577

QY 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTrpLeuTrpValLysSerValLys 242
 Db 578 GAGGCAATGGCAATATTCGCCCTCGCAGTTTCACCTACTTCTCTGGTGGTGGTGGTGGTGGT 637
 QY 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMetValSerAla 262
 Db 638 ACTGGATCCGTTCTGGTGGCGCGCGCTTTGCTACTTCTACTTCTACTTCTACTTCTACTT 697
 QY 263 TrpGlyGlyTrpValPheIleIleLeuLeuIleProLeuHisValPheValLeuLeu 282
 Db 698 TGGGTGGTACGTTCTCATCAACCTGATACCCCTGCACGCTCTTCTGCTACTTCTGCTACT 757
 QY 283 MetGlnArgTrpSerLysArgValTyrIleAlaTrpSerThrPheTrpIleValGlyLeu 302
 Db 758 ATGGCAGGTACTTCGGCGGCTGTGTGACCACTACACGACCTTCTACTTCTACTTCTACTT 817
 QY 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 Db 818 CTGTTCTCCATGAGATCCCTTCGTGGGATTCACACCGATACGACCACTGACACACATG 877
 QY 323 AlaAlaAlaGlyValPheAlaLeuGlnAlaTrpAlaPheLeuGlnTrpLeuArgAsp 342
 Db 878 GCTGGCTGGAGTGTGTGCTCTTATGGCGGTGGCCACCTTGGCCATTTGCAGTCC 937
 QY 343 ArgLeuThrLysGlnPheGlnThrLeuPheLeuGlyValSerIleAlaAlaGly 362
 Db 938 GTGCTGTCGGCAACGAGTTCGGAGCTGTTTCATCGTCGGCGGATTCGCTGGTGGCGGTT 997
 QY 363 AlaValPheLeuSerValIleTrpLeuThrTrpThrGlyTrpIleAlaProTrpSerGly 382
 Db 998 GCGCTCTTGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1057
 QY 383 ArgPheTrpSerLeuTrpAspThrGlyTrpAlaLysIleHisIleProIleAlaSer 402
 Db 1058 CGTCTTACTCGTGGGATCTGGCTACGCAAGATCCACATTCCTCCATTCATTCATTCATTC 1117
 QY 403 ValSerGluHisGlnProThrTrpValSerPhePhePhePhePhePhePhePhePhePhe 422
 Db 1118 GTGTCGGAGCATCAGCCACCACTTGGTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1177
 QY 423 CysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPhe 442
 Db 1178 TGGCCCTTCCATGGAGGTGGTACTGCATCAGCAGATCAACGACGAGCGGCTTTTC 1237
 QY 443 ValAlaLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetValArgLeuMetLeu 462
 Db 1238 GTGCTGCTGTACGCCATCAGTGGGTTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1297
 QY 463 ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482
 Db 1298 ACCCTACCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1357
 QY 483 TyrLeuGlyAspAsp---MetLysArgGluAsnProProValGluAspSerSerAspGlu 501
 Db 1358 TTCCTGCAAGAGGATTCGTCTAAGCGAATGGGCACAGCCATTAAGCGGCGACCCAGGAGTG 1417
 QY 502 Asp-----AspLysArgAsnGlnGlyAsnLeuTrpAspLysAlaGlyLysValArg 518
 Db 1418 GATGAAGCTGAGGATTCATTTGAGAGAGAGAGCGCTGTACGACAGGCTGGCAGCTG--- 1474
 QY 519 LysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsnIleLysSer 537
 Db 1475 AAGCATCGTACTAAGCATGATGCCAGCAGGATCTGGCGTCAGCTCCAACTCCAACTGAAGAGT 1534
 QY 538 IleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
 Db 1535 ATTGTTATTTGGCCGTTCTAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1594
 QY 558 ThrSerAsnAlaTrpSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
 Db 1595 ACCAGCAATGCTACTACTGCTCCCTCCATTTGCTTGGCTTTTCCACACAGTCAAGATGGA 1654

Db 1805 ATATTAGATGATTTTACGAGAGCGTACTTTTGGCTGAGACAAACACGATGACACGCGC 1864
 Qy 600 ArgValMetSerTTPtTpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThr 619
 Db 1865 CGGGTCATGTCGTGGTGGACTACGGTATCAGATTGCTGGCATGGCCACAGACGACCT 1924
 Qy 620 LeuValAspAsnAsnThrTTPtTpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSer 639
 Db 1925 CTGGTGGATACACACCTGGAACACAGCCACATCGCACTGGTCGGAAGCTATGTCT 1984
 Qy 640 SerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuVal 659
 Db 1985 TCCATGAAAGCGCGCTATAAATCATGAGTCCCTTGATGCTGATGATGTGTGT 2044
 Qy 660 IlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetVal 679
 Db 2045 ATTTTCGAGAGGATGATGGCTATTCGCGGAGCATATCAACAAGTTCCTCTGGATGGTC 2104
 Qy 680 ArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGln 699
 Db 2105 AGGATAGCTGAAGGAGCATCCCAAGACATCCGGAAGGTGACTATTTCACCCAGCAG 2164
 Qy 700 GlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLys 719
 Db 2165 GGAGAGTTCGAGTAGACAAAGCTGGGTCTCTACTCTGTATAACTGCCTTATGTATAAA 2224
 Qy 720 MetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAsp 739
 Db 2225 ATGTCATACTACAGATTGGAGAAATGACAGTAGATTTCGCATCCCCAGGCTTTGAC 2284
 Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAla 759
 Db 2285 CGAACACGTAATGCTGAGATTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 2344
 Qy 760 PheThrSerGluHisTTPtTpLeuValArgIleTyrLysValLysAlaProAspAsnArgGlu 779
 Db 2345 TTTACATCAGACGACTGGCTGTGACAGATATATAAGTGAAGCACTGACACAGGGAG 2404
 Qy 780 ThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLys 799
 Db 2405 ACATAGTGTCAAACTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAG 2464
 Qy 800 LysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLys 819
 Db 2465 AAGACTACTAAGAGAGCGGTGCTAGCTGTTAAAAATAGCTAGCTGTTAAGAAGGCAAG 2524
 Qy 820 LysIleSerLysLysThrVal 826
 Db 2525 AAGACCTCTAAGAGACTGTT 2545
 RESULT 5
 AF132552 2417 bp mRNA linear INV 12-FEB-2003
 LOCUS Drosophila melanogaster GNC1038 full insert cDNA.
 DEFINITION AF132552
 ACCESSION AF132552
 VERSION AF132552.1 GI:4689327
 KEYWORDS FLI CDNA.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 2417)
 Tsang, G., Broksstein, P., Frise, E., Harvey, D., Evans-Holm, M.,
 Lewis, S.E., Suh, C. and Rubin, G.M.
 Direct Submission
 Submitted (02-MAR-1999) Berkeley Drosophila Genome Project,
 University of California Berkeley, 539 Life Sciences Addition
 #3200, Berkeley, CA 94720, USA
 JOURNAL 2 (bases 1 to 2417)
 REFERENCE 2 (bases 1 to 2417)
 AUTHORS Stapleton, M., Broksstein, P., Hong, L., Agbayani, A., Carlson, J.,
 Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E.,

George, R., Gonzalez, M., Guarin, H., Krommiller, B., Li, P., Liao, G.,
 Miranda, A., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
 Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M.
 and Celniker, S.
 Direct Submission
 Submitted (12-FEB-2003) Berkeley Drosophila Genome Project,
 Lawrence Berkeley National Laboratory, One Cyclotron Road,
 Berkeley, CA 94720, USA
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory
 Berkeley, CA 94720
 This clone was sequenced as part of a high-throughput process to
 sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
 Science 2000). The sequence has been subjected to integrity checks
 for sequence accuracy, presence of a polyA tail and contiguity
 within 100 kb in the genome. Thus we believe the sequence to
 reflect accurately this particular cDNA clone. However, there are
 artifacts associated with the generation of cDNA clones that may
 have not been detected in our initial analyses such as internal
 priming, priming from contaminating genomic DNA, retained introns
 due to reverse transcription of unspliced precursor RNAs, and
 reverse transcriptase errors that result in single base changes.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our web site
 (http://fruitfly.berkeley.edu) or send email to
 cdna@fruitfly.berkeley.edu.

FEATURES

1..2417
 Location/Qualifiers

gene
 <1..2417
 /gene="OstStt3"
 /db_xref="FLYBASE:FBgn0011336"
 CDS
 <1..2395
 /gene="OstStt3"
 /notes="Longest ORF"
 /codon_start=2
 /product="Gm01938p"
 /protein_id="AAD27851.2"
 /db_xref="GI:28317200"
 /db_xref="FLYBASE:FBgn0011336"
 /translation="ISEECVAFVFIQVFNRTVGMNMDNRTPKMLNSKVAGVSSLIIF
 AILLIAWAFSSRIFAVIRFVFIHEPDPENYRATAYMVQNGWYNLNNPDERAWY
 PLGRIVGTVPLGLMITSGIHLHLVNLPHIRDCVFLAIPSGLTISTYLTATK
 ELWSACAGI PAASFAIIVPGYISRSVAGSYDNEGAI PAI PAI PAI PAI PAI PAI PAI
 SAAALSYFTMSAWGGYVFIINLPLHVFVLLIMGRISPRLLISYFYILGLLFSM
 QIBFVGQPIRTSEENAAAGVFLMLAVATLHLOSILRSNFRKLFIVGGLLVGVG
 FVAVVLTMLGVVAPWSGRFYSMTDGTAKIHIPIASVSEHQPTTFSPFDHLILY
 DFLQEDSSKRMGTAI SAATEVDEADSI EKKTLYDKAGKLKRTKHDAQODTGVSSN
 CAPFVGWYCIQINDERYVVLVYIAISAVYFAGVNVRLMLTLPVVCMLAGVAFSGLL
 LKSVILAVLMLLMFAVCTVWTNSYSSPISVLAHNSODGSRNILDDPREAVYL
 SONTADARVMSWDYGYOIAGMNNTTLDNNTNNSHIALVGMKSTEEKSYEIM
 TSLDYVLVIFSGVIGYSGDDINKFLMWRIAEGEHPKDIKESDYFTDRGEFRVDAE
 GAPALLNLMYLSYRFGELKLDVRGPGYDRTNAVIGNKDFDLTLEBAYTTEHW
 LVRIYRVKPHPEFRPSLTKERTTIPANFISRKNSKRRKGYIRNPVVVKGKRLTK"

ORIGIN

Alignment Scores:

Pred. No.: 2,14e-227 Length: 2417
 Score: 2950.00 Matches: 557
 Percent Similarity: 83.79% Conservative: 89
 Best Local Similarity: 72.24% Mismatches: 113
 Query Match: 72.03% Indels: 12
 DB: 3 Gaps: 7

US-10-028-384-2 (1-826) x AF132552 (1-2417)

Qy 63 SerGlnProAlaGlyTTPGlnSerLeuSerPheThrIleLeuPheLeuAlaThrPleu 82

Db 98 AGCAAGTGGTGGCTACAGCAGCCTAATCACCTTCGCCATCTGCTAATCGCTGGCTG 157

VERSION AX799084.1 GI:37605059
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Perreault, C. and McBride, K.
 TITLE Mammalian SAMP protein, gene sequence and uses thereof in cancer therapy
 JOURNAL Patent: WO 03054008-A 3 03-JUL-2003;
 COMPATIGENE Inc. (CA)
 FEATURES
 source
 1..2710
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 ORIGIN
 Alignment Scores:
 Score: 0 Length: 2710
 Pred. No.: 4263.50 Matches: 805
 Percent Similarity: 97.70% Conservative: 3
 Best Local Similarity: 97.34% Mismatches: 14
 Query Match: 96.88% Indels: 5
 DB: 6 Gaps: 2
 US-10-028-384-2 (1-826) x AX799084 (1-2710)
 QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTrp 20
 Db 77 ATGGCGGAGCCCTCGGGCCCGGAGAGACAGACAGAGTCGCTCCCAACTCGTCCCGTGG 136
 QY 21 SerGlyLeuMetAlaLeuGluAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 Db 137 AGCGGCTCATGGCTCTGGGGAAGAGCGGCGACCGGACCATCGGCGCGGAGCCAGGC 196
 QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAla---ProAlaGlyLeuSer 59
 Db 137 GCGTCCAGGGCGGGGGG-----CCGAAGCGGGGGGGGGGGGGGGGGTGTCC 244
 QY 60 GlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuSerPheThrIleLeuPheLeu 79
 Db 245 GGGGGCTGTGCGAGCGGGCGGGTGGCAGTGTGCTCTCTCTCTCTCTCTCTCTCTCT 304
 QY 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIle 99
 Db 305 GCCTGGCTGGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAGAGCATATC 364
 QY 100 HisGluPheAspProTrpPheAsnTrpArgSerThrHisIleuAlaSerHisGlyPhe 119
 Db 365 CACAGTTCGACCGCGTGTAACTATGATACACACATCATCTTCATCTCATGGATTC 424
 QY 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTrpLeuGlyArgIleValGly 139
 Db 425 TATCAGTTCCTAAATGGTTGTATGATGAAGAGCATGTGTACCCACTGGGAAGATAGTGGT 484
 QY 140 GlyThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
 Db 485 GGCACCGTTTACCCAGGGTGTATGATTAACAGCTGGCGCTTATTCATTTGGATTTAATACA 544
 QY 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
 Db 545 TTGAACATAACAGTTCATGAAGAGATGTGTGTATTCCTTGCACCAACTTTTACCGGC 604
 QY 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199
 Db 605 CTTACATCCATATCATGTTCTCTGCTAAGTGTAGAGACTGTGGNACCAAGAGAGAGACTT 664
 QY 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
 Db 665 CTAGCTGCTTCTCTTCTTCTATCGTACCGAGGGGTATATATCTCGGTTCAGTGGCGGATCC 724

220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpVallys 239
 Db 725 TTTGATAATGAAGGCATTGCCATTTTGGCTTCAGTTCACTTACTTATGGGTAAAG 784
 QY 240 SerValysThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTrpMet 259
 Db 785 TCTGTGAAGACCGGGTCTGTGTCTGGCAATGTGTCTGTCTGTCTATATTTCTACATG 844
 QY 260 ValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
 Db 845 GTCTCTGGTGGGAGGTATGTGTTCATCATCACTCACTCCCTCTCCATGTGTGTGG 904
 QY 280 LeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTrpIle 299
 Db 905 TTGCTGTGTATGACAGAGGTACAGCAAGAGAGTCTATAGCATATAGCACATTGTACAT 964
 QY 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
 Db 965 GTGGGTTTAAATATATTCATGCAGATACCTTTTGTGGATTTTCAGCCANTCAGAACAGC 1024
 QY 320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTrp 339
 Db 1025 GAGCACATGGCAGCTGCAGGTGTCTTTGGCGCTGCTGCAGGCTTACGCTTTTTCAGTAT 1084
 QY 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
 Db 1085 CTGAGAGACCGGTTGACAAAACAGGAGTTCAGAGCCCTTTCTTTTGGGTGTCTCATA 1144
 QY 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrGlyTyrIleAlaPro 379
 Db 1145 GCTGCAGCGGCTGTGTCTCTTGTGTCTATCTATCTGACATACACAGGTATATTTGCACA 1204
 QY 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399
 Db 1205 TGAGTGGCGAGTTTATTTACTATGGGATACCTGGGTATGCAAAATACACATTTCCAAT 1264
 QY 400 IleAlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHis 419
 Db 1265 ATTGCATCATGTGTCTGAACATCAGCCTCAGACATGGGTGTCTTTCTTTGATCTACAT 1324
 QY 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGlu 439
 Db 1325 ATCTTTGTATGTACCTTCCACAGCGCTATGGTTCTGCATCAAAATATATCAACATGAA 1384
 QY 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
 Db 1385 AGAGTATTTGTCTGTATGCGATCAGTGTGTGTACTTTTGGCGGAGTGTGGTGGG 1444
 QY 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
 Db 1445 CTGATGCTGATCTGACCCCGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1504
 QY 480 PheGluHisTyrLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSer 499
 Db 1505 TTTGAGCACTATTGGGGGATGACATGAAAGGGGAAACCCACCTGTGGAGAGCAGCAGT 1564
 QY 500 AspGluAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLys 519
 Db 1565 GATGAGATGACAAAGAAACCCAGGAACCTTGTATGACAGGCGAGTAAAGTGGAGAG 1624
 QY 520 HisAlaThrGluGlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleVal 539
 Db 1625 CATGTGACAGAGCAAGAGAAACCTGAGAGGGGCTGGGGCCCCCAACATCAAAAGCATTTGT 1684
 QY 540 ThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSer 559
 Db 1685 ACCATGCTGATGCTCATGCTCTCTGATGTGTGGGGTTCCTGCTGCTGCTGCTGCTGCT 1744
 QY 560 AsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsn 579
 Db 1745 AAGCGCTACTCAGTCCAGTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1804
 QY 580 IleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAla 599

QY 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
Db 774 GGCACCGTTTACCCAGGGTTGATATAACAGCTGGCTTATTCATTGGATTATAATACA 833
QY 160 LeuAsnIleThrValHisIleArgAspValCysValPheIleuAlaProThrPheSerGly 179
Db 834 TTGAACATAAAGAGTTTCAATAGAGATGTGTGTATTCCTTGCACCAACTTTTAGCGGC 893
QY 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199
Db 894 CTTACATCCATATCTACGTTCTCTGCTAACTAGAGAACTGTGGAACCAAGAGACGAGCTT 953
QY 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
Db 954 CTAGCTGCTTCTTCATCTGCTATGCTACAGAGGTACATATCTCGGTGAGTGGCGATCC 1013
QY 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTrpLeuTrpValLys 239
Db 1014 TTTGATAATGAAGCAATGGCAATTTTGGCTTCAGTTCACTTACTTATGGGTAAAG 1073
QY 240 SerValIleThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMet 259
Db 1074 TCTGTGAAGACCGGCTCTGTGTTCTGGACAAATGTCTGCTTGTTCATATTTTACATG 1133
QY 260 ValSerAlaTrpGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
Db 1134 GTCTCTGGTGGGAGGTTATGTGTTTCATCATCAACCTCATCCCTCTCCATGTGTGTTG 1193
QY 280 LeuLeuLeuMetGlnArgTyrSerValPheValTyrIleAlaTyrSerThrPheTyrIle 299
Db 1194 TTGCTGCTGATGACAGGTACAGCAAGAGTCTTACATGATATAGCACTTTTCTACATT 1253
QY 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
Db 1254 GTGGTGTAAATATTATTCATGACATACCTTTTGTGGATTTCAGCAATTCAGAACAGC 1313
QY 320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
Db 1314 GAGCATGTGGCAGCTGCGAGGTCTTTGGCTGCTGCAAGCTTACCCTTTTGTGCAAGTAT 1373
QY 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
Db 1374 CTGAGAGACCGGTTGACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGCTCACATA 1433
QY 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
Db 1434 GCTGAGCGGCTGTGTTCTTGTGTCATCTATCTGACATACAGGTTATATGACCA 1493
QY 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399
Db 1494 TGGAGTGGCAGGTTTATTTCACTATGGATACCTGGGTATGCAAAAATACACATTCCAATT 1553
QY 400 IleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHis 419
Db 1554 ATTGCATCAGTGTCTGAACATCACCCCTACGACATGGGTGCTTCTTCTTGTATCTACAT 1613
QY 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGlu 439
Db 1614 ATCTTGTATGTACCTTCCAGCAGGCTATGGTTCTGTCATCAAAAATACCAAGATGAA 1673
QY 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
Db 1674 AGAGTATTGTGCTCTGTATGCGATCAGTGTCTTACTTTCGGGAGTGTGGTGGCG 1733
QY 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
Db 1734 CTGATGCTGACTCTGACCCCGGTCGTCTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1793
QY 480 PheGluHisTyrLeuGlyAspAspMetIleArgGluAsnProProValGluAspSerSer 499
Db 1794 TTTGAGCACTATTGTGGGGGATGACATGAAGAAGGGAACCCACCTGTGGAGGACAGCT 1853

QY 500 AspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLys 519
Db 1854 GATCAGAGTGCACAAAAGAAACCCAGGAAACTTGTGTATGCAAGGCGAGTAAAGTGGGAAG 1913
QY 520 HisAlaThrGluGlnGlyThrGluGluGlyLeuGlyProAsnIleLysSerIleVal 539
Db 1914 CATGTGACAGAGCAAGAGAAACCTGAGAGGGCTTGGGCCCCACATCAAAAGCATTTGTG 1973
QY 540 ThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTyrTrpValThrSer 559
Db 1974 ACCATGCTGATGCTCATGCTCTCTGATGTTTTCGGGCTCCATGTCACGTGGGTGTCACAGC 2033
QY 560 AsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsn 579
Db 2034 AAGCCTACTCTCAGTCCAGTGTGGTCTTCTCTTACATCATGATGTTACCAAGGAT 2093
QY 580 IleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAla 599
Db 2094 ATATTAGATGATTTTAGAAGAGCTACTTTTGGCTGGAGCAAAACACGAGTGAACACGCC 2153
QY 600 ArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThr 619
Db 2154 CGGCTCATGCTGCTGGTGGACTAGGCTATCAGATTGCTGGCATGCCCAACAGGACCACT 2213
QY 620 LeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSer 639
Db 2214 CTGCTGGATAAACCAACACCTGGAAACACAGCCATCGCACCTGGTCGGAAGAGTATGCT 2273
QY 640 SerAsnGluThrAlaAlaTyrIleMetArgThrLeuAspValAspTyrValLeuVal 659
Db 2274 TCCATGAACCGCCGCTTATAAATCATGAGGTCCCTTGATGTCGATTGTTGTT 2333
QY 660 IlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetVal 679
Db 2334 ATTTTCGGAGAGTGTGCTTATTCGGGGAGCATATCAACAGTTCTCTCTGGAATGTC 2393
QY 680 ArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGln 699
Db 2394 AGATAGCTGAGGGGAGCATCCCAAGACATCCGGGAAGGTGACTATTTCACCCAGCAG 2453
QY 700 GlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLys 719
Db 2454 GGAGAGTCCGAGTAGAACAAAGCTGGTCTCTACTCTGTTAACTGCTTATGATATAA 2513
QY 720 MetSerTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAsp 739
Db 2514 ATGTCATCTACAGATTTCGAGAAATGCGAGCTAGATTTCGCACCTCCCGAGCTTTGAC 2573
QY 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAla 759
Db 2574 CGAACACGTAATGCTGAGATTGGAAATAAAGACATTAATTCACGCAATTTGAGGAAGCT 2633
QY 760 PheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGlu 779
Db 2634 TTATCATCAGAGCAGCTGGCTGTGTCAGATATATAAAGTAAAGACCTTCACACAGGAG 2693
QY 780 ThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLys 799
Db 2694 ACCTAGGTCAACAACCTCGAGTCACCAACATCGTCCCAAAACAGAAGTATTGTGCAAG 2753
QY 800 LysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLys 819
Db 2754 AAGCTACTTAAAGGAAGCGTGGCTACGTTAAATATAGCTAGTGTTTAAGNAAGGCAAG 2813
QY 820 LysIleSerLysLysThrVal 826
Db 2814 AAGACCTCTAAGAAGACTGTT 2834

RESULT 4

AX799084

LOCUS

DEFINITION

ACCESSION

AX799084 2710 bp mRNA linear PAT 06-OCT-2003

Sequence 3 from Patent WO03054008.

AX799084


```

BC052433.1 GI:30851501
MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Datchenko, L., Nardusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, A.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, K.J., Mullen, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shvachenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smalls, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22384257
1 (bases 1 to 4236)
2 (bases 1 to 4236)
Submitted (15-MAY-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-4590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas L. Casavant.
Web site: http://genome.uiowa.edu
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldi, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,
Fishler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J.,
Casavant, T., Soares, M.B.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
1. .4236
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:64679 IMAGE:6837097"
/tissue_type="Brain, enriched mouse brain 12.5dpc"
/clone_lib="NIH-BMAP_F10"
/lab_host="DH10B"
/notes="vector: pYX-ASC"
1. .4236
/gene="1300006C19Rik"
/notes="synonym: Simp"
/db_xref="LocusID:68292"

```

```

/db_xref="MGI:1915542"
366_ .2837
/gene="1300006C19Rik"
/codon_start=1
/product="1300006C19Rik protein"
/protein_id="AAH52433.1"
/db_xref="GI:30851502"
/db_xref="LocusID:68292"
/db_xref="MGI:1915542"
/translacion="MAEPSPKSHKSLNSPWSGLMAGNSRHGHGPGTQASRA
APKCPGAPAGSGLSQACWQSLSTFLILAMLAGPSSLEAVIESIIEFDPW
FYRSTHLDASHGFYELNWFDERAWYELGVGTVPYGLMILAGIHLINLINT
VHIRDVCFLAFTFSLSTISFLDLRELWAGGLAACFIALIPYIGRSVAGSD
NGAIFALQFTYILWKSIVKTSVFTMCCLSYFYFVMSWAGGYFIILPHVEF
LLMLQSKRYVYIAYSTFYIVGLILSMQIPFYFQPISTSEHMAAGVALLQYAFV
QILRLRLKQEPQTLFFGLSLAGAVLSVILYITGYIAPMSGRYSVLDWTGYAKI
HIPIASVSEHPTTWSVFFDFDLHVLCTFFAGLVFCIKINDEKRVAVLVAIVY
AGVYVLMILTLTPVCMLSAFAFNVFHYLGDMDKRENPEVSDSDKRNPGNLY
DQAGKRVHVTQEPKPEGLGPKISIVTLMMLMMAFAVACTWVTSNAYSPPSVL
ASYHNDGTRNILDPRFAYFWLRQNDIHRVMSWWDYGOIAGMARTILVDNNTWN
NSHIALVGMSSNETAAYKIMRSLDYLIVFGVIGYSGDDINKFLMWRVARE
HPKIDREGDYFTQQGEFRVDKAGSPTLLNLMYKSYRFGEMQDPTPPGDRRN
AETGNKDIKPHLEBAFTSEHVLVRIYKVPADNRETILGHKPRVTNIVPKQKLSKKT
TKRKGIVKVKLVFKKKTSTKTV"
567_ .2537
/misc_feature
/gene="1300006C19Rik"
/notes="SRT3; Region: Oligosaccharyl transferase SRT3
subunit. This family consists of the oligosaccharyl
transferase SRT3 subunit and related proteins. The SRT3
subunit is part of the oligosaccharyl transferase (Oase)
complex of proteins and is required for its activity.
Oase transfers a lipid-linked core-oligosaccharide to
selected asparagine residues in the ER"
/db_xref="CDD:pfam02516"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4236
Score: 4269.50 Matches: 806
Percent Similarity: 97.92% Conservative: 3
Best Local Similarity: 97.46% Mismatches: 13
Query Match: 97.01% Indels: 5
DB: 10 Gaps: 2
US-10-028-384-2 (1-826) x BC052433 (1-4236)
Qy 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTirp 20
Dy 366 ATGGCGGAGCCCTCGGCCCGGAGAGCAAGCACAAGTGTCTCCCTCACTCGTCCCGGTGG 425
Qy 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisHisGlyProGlyAlaGlnCys 40
Dy 426 AGCGGCTCATGCTCTCGGGAAACAGCCGCCACGAGCCACCATGGCGCGGACCCAGAGC 485
Qy 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProlLysProAla---ProAlaGlyLeuSer 59
Dy 486 GCGTCCAGCGCGCGCGCGG-----CGAAGCGCGCGCGCGCGCGCGCGCGCTGCC 533
Qy 60 GlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrLeuLeuPheLeu 79
Dy 534 GGGGGCTTGTTCGACGCGCGCGGGTGGCAGTGTGTCTCTCTTCATCATCTCTCTCTG 593
Qy 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIle 99
Dy 594 GCGTGGCTGCGCGGCTTCAGCTCGCGCTCTTCGCGGTTCATCCGCTTCGAGAGCATCATC 653
Qy 100 HisGluPheAspProTirpPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPhe 119
Dy 654 CACGAGTTCGACCGCGGTGTTAACTATAGATCAACATCATCTTGTGATCTCATGAGTTC 713
Qy 120 TyrGluPheLeuAsnTirpPheAspGluArgAlaTirpTyrProLeuGlyValGly 139
Dy 714 TATGAGTTCTTAATTTGGTTTGTGTAAGAGAGCATGGTACCCACCTGGGAAGATAGTGGT 773

```


141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
421 ACTGTTTACCCAGGGTTGATCAACCGCTGGCTTATTCATTGGATTTTAATACATGG 480
161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
481 AACATAACTGTTTCATAGAAGAGAGCTATGCTGCTTCCTGCACCAACTTTTAGCGGCTT 540
181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
541 ACATCTATATCTACTTCTCTCTTACAGAGAACTTTGGAACCAAGAGGACGACTTTTA 600
201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
601 GCTGCTGTTGTTTATGCTATGTTACAGGCTACATATCTCGGTCACTAGCTGATCTCTT 660
221 AspaGlnGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrIleTrpValIysSer 240
661 GATAATGAAGGCAATGCTATTTTGGCACTTCAGTTCACATACATTTATTTGGGTAATCT 720
241 ValIysThrGlySerValPheTrpThrMetCysCysCysLeuSerTyrPheTyrMetVal 260
721 GTAAACCTGGCTCAGTTTGTGACAACTGCTGCTCTATCTTATTTCTATATGCTC 780
261 SerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
781 TCTGCTTGGGTGGTATGATTTATCATCAATCTTATCCACTGCATGATTTTGGTGG 840
281 LeuLeuMetGlnArgTyrSerIleArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
841 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATG 900
301 GlyIleIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
901 GGTTTAAATATCATCAATGCAATACCTTTTGTGGGATTCAGCAATACAGAAAGTGA 960
321 HisMetAlaAlaAlaGlyValPheAlaLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
961 CACATGGCAGCTGCAGGTGCTTTTGGATTCGCAAGCTTATGCTTCTTTCAGTATCTG 1020
341 ArgAspArgLeuThrLysGlnGlyPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
1021 AGAGCCGATTAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGATCAGTACCT 1080
361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrp 380
1081 GCAGGTGCTGTTCTTCTAGTGTCTATCTATTTGACTTATACAGGTTACATTCGACCATG 1140
381 SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIle 400
1141 AGTGCAGGTTTATCTATCTGGGATCTGGGTATGCAAAATACACATTCCAATATT 1200
401 AlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHisIle 420
1201 GCATCAGTGTCTGAGCAACCTACAGTGGGTGCTTCTTCTTCTGATCTACATATT 1260
421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
1261 CTGTATGATCTTCCAGCAGGCTTTGGTCTGCAATCAAAATATCAACGATGAAGA 1320
441 ValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeu 460
1321 GTATTGTTGCTCTATATGCAATCAGTGTCTCTACTTCTGCTGAGTGTGGTGCAGCTG 1380
461 MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
1381 ATGTTGACTTTGACTCCAGTCTGCTGATGCTGCTCAATGCTCTTCTCAATGTTT 1440
481 GluHisTyrLeuGlyAspMetLysArgGluAsnProProValGluAspSerSerAsp 500
1441 GAGCACTATTTGGGGATGATCAATGAAAGGAAATCCACCTGTGGAGGACAGCAGTAT 1500
501 GluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHis 520

1501 GAGGATGACAAAAGAACCAAGGAATTTGTATGATAAGCGAGGTAAAGTGAGGAAACAT 1560
521 AlaThrGluGlnGlyThrGluGlyLeuGlyProAsnIleLysSerIleValThr 540
1561 GCAACTGAACAGGAAACAACTGAAGAGGATTAGGCCCTAATATATAAAGCATTTGTACC 1620
541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
1621 ATGTTGATGCTGATGCTATTTGATGATTTGCTGCTCCACTGTACCTGGGTCAACAGCAAT 1680
561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
1681 GCCTACTAGTCCAAAGTGTAGTCTCTGGCTCATCAATCATGATGCGCACCAAGGAATATC 1740
581 LeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
1741 TTAGATGATTTTGAAGAAGCTTACTTTTGGCTTAAGCAAAATACAGATGAACATGACAGA 1800
601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
1801 GTAATGCTTGTGGGATTTATGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG 1860
621 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
1861 GTGGATTAATAACACCTGGGAATACACCCACATAGCACCTGTGGGAAAGACTATGTTCT 1920
641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIle 660
1921 AATGAAACAGCAGCTTATAAATCATAGGACTCTAGATGTAGATATGTTTGTGTTATT 1980
661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
1981 TTTGGAGGGTTATTTGGCTATTTCTGTTGATGATATCAACAATTTCTCTGGATGTTAGG 2040
681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
2041 ATAGCTGAGGAGAACATCCAAAGACATTCGGGAAAGTACTATTTTACCCACAGGGA 2100
701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
2101 GAATTCGTTAGCAAAAGCAGGATCCCTACTTTTGTGTAATGCTTATGATATAAATG 2160
721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArg 740
2161 TCATCTACAGATTTGGAGAAATGTCAGCTGGATTTTTCGTACACCCCGAGTTTTCACCGA 2220
741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPhe 760
2221 ACAGTAACTCTGAGATTGGAATAAGGACATTAATTCAAACATTTTGAAGAAGCCTTT 2280
761 ThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
2281 ACATCAGAACACTGGCTTGTAGGATATATAAGTAAAGACACCTGATAACAGGAGACA 2340
781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
2341 TTAGATCACAACACTCGAGTCACCAACATTTTCCCAAAACAGAGATATTGTTCAAGAAG 2400
801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysGlyLysLys 820
2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAAGCTGGTTTAAAGAAAGGCAAGAAA 2460
821 IleSerLysLysThrVal 826
2461 ATAATCTNAGAGACTGTT 2478

RESULT 3

BC052433

LOCUS

DEFINITION Mus musculus RIKEN cDNA 1300006C19 gene, mRNA (cdna clone MGC:64679 IMAGE:6837097), complete cds.

ACCESSION BC052433

BC052433 4236 bp mRNA linear ROD 12-NOV-2003

AGGAAPKAPAGLSGLSOPAGWQSLLSTLILFLAWLAGFSRLFAVIRPESIIHEF
DPWFNYSRTHLSHSGYEFELNFWDERAWPLGRIYGGTVIPGLMITAGLIHWLNTL
NITVHRDVCVFLAPTSGSLTSISFTLLRELNQAGLAAFCIAIPVYSRSVAG
SFDNEGAIYFALQYLYLWVKSFTVCCSLYSFTVMSAWGQYFIINLIPLH
VFVLLLMQRYSTKRYIAYSTFYIVGLILSMQIPFVGFPQIRTSERHMAAGFALLQAY
AFQYLRDRITKQBFQTFELGSLAAGAVLSVILYITGYIAPWSGRFYSLMDTGY
AKIHIPIASVSEHPTTWYSFFDLHILVCTPAGLWFCIKNINDERFVVALYISA
VYPAGVMRLMLTLTPVCMLSALAFSPNVFHYLGDMMKRENPVEDSDEDDKRNQ
NLVDKAGKVRKHATEQKTEEGLPNPKSIVTMLMLMLMPFAVCTWISNAYSSES
VFLASYNHEDGTRNLLDFREAYFWLQNTDEHARVMSWDYGOIAGMANRITLVDNN
TWNHSHLVGKAMSNETHAAYKIMRTLDYDVLVIFGGVYIGYGGDDINKFLMWYIA
EGEHPKDIRESDYFTPOGEFRVDPKAGSPILLNCLMYKMSYREFGEMQDPRPPGDR
TRNABIGNKDKFKHLEAFTSEHVLVRIYKVKAPDNRETLDHKPRVTNIPFKQKLYS
KKTITKRGYIKNKLVPKGGKISKXTV"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2481
Score: 4401.00 Matches: 826
Percent Similarity: 100.00% Conservations: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-028-384-2 (1-826) x AX799082 (1-2481)

Qy	1	MetAlaGluProSerAlaProGluSerLysHisLysSerLeuAsnSerSerProTrrp	20
Db	1	ATGGCGAGCCCTCGGCGCGGAGAGCAACAGTCGTCCTCACTCGTCCCGTGG	60
Qy	21	SerGlyLeuMetAlaLeuGlyAenSerArgHisGlyHisGlyProGlyAlaGlnCys	40
Db	61	AGTGGCCCTATGGCCCTGGGAACAGCGGCGAGCCGCGGCGGCGGCGGCGGCGG	120
Qy	41	AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSergly	60
Db	121	GGCGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	180
Qy	61	GlyLeuSerGlnProAlaGlyTrrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla	80
Db	181	GGGCTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	240
Qy	81	TrrpLeuAlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleHis	100
Db	241	TGGCTTGGCGGCTTACGTGCGGCTTTCGCGGCTATCGGCTTGAAGAGCATATCCAC	300
Qy	101	GluPheAspProTrrpPheAsnTrrpArgSerThrHisHisLeuAlaSerHisGlyPheTyr	120
Db	301	GAGTTCGACCGGTGGTAACTATAGATCAACACATCATCTTGCATCTCATGGTTCAT	360
Qy	121	GluPheLeuAsnTrrpPheAspGluArgAlaTrrpTrrpProLeuGlyArgIleValGlyGly	140
Db	361	GAATTTTAAATTTGGTTTATGAAGAGCATGTATCCACTAGGAGAAATAGTAGTGGT	420
Qy	141	ThrValTrrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrrpIleLeuAsnThrLeu	160
Db	421	ACTGTTTACCCAGGTTGATGATAACCGTGCGCTTATTCATGGATTTTAAATACATTG	480
Qy	161	AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu	180
Db	481	ACATTAACGTGTACATAAGACGATGTGTGTCTTCCCTTGCACCAACTTTTAGCGGCTT	540
Qy	181	ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrrpAsnGlnGlyValaGlyLeuLeu	200
Db	541	ACATCATATCTACTTTCGCTTACAGAGAACTTTTGAACCAAGGAGGAGGAGCTTTTAA	600
Qy	201	AlaAlaCysPheIleAlaIleValProGlyTrrpIleSerArgSerValAlaGlySerPhe	220
Db	601	GCTGCTGTGTTTATGCTATTACAGGCTACATATCTCGGTACGTAGTGGATCCCTTT	660
Qy	221	AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrrpTrrpValIlySer	240
Db	661	GATATGAGGCAATGCTATTTTTCGACTTCAGTTCACATATCTATTTATGGGTAAATCT	720

Qy	241	ValLysThrGlySerValPheTrrpThrMetCysCysLeuSerTrrpPheTrrpMetVal	260
Db	721	GTAATAAACTGGTCACTTTTGGACATGCTGCTCTTATCTTCTATATGGTC	780
Qy	261	SerAlaTrrpGlyGlyTrrpValPheIleIleAsnLeuIleProLeuHisValPheValLeu	280
Db	781	TCGTGCTGGGGGGTATGATTTATCATCAATCTTATTCACCTGCATGATTTTGTGTG	840
Qy	281	LeuLeuMetGlnArgTrrpSerLysArgValTrrpIleAlaTrrpSerThrPheTrrpIleVal	300
Db	841	TTACTGATGCAGAGATACAGCAAAAGAGCTTACATAGCATATAGCATTTCTCATTTGTG	900
Qy	301	GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu	320
Db	901	GGTTTAATATTATCAATGCAGATACCTTTTGTGGATTTCCAGCCAATCAGAACAGTAA	960
Qy	321	HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTrrpAlaPheLeuGlnTrrpLeu	340
Db	961	CACATGGCGGCTGAGGCTGCTTTTGCATGCTGCAAGCTTATGCTTTCTTGCATATCTG	1020
Qy	341	ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla	360
Db	1021	AGAGACCGATTAAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGATCATCTAGCT	1080
Qy	361	AlaGlyAlaValPheLeuSerValIleTrrpLeuThrTrrpGlyTrrpIleAlaProTrrp	380
Db	1081	CGAGTGTGCTGCTTCTTGTAGTCTATTTTGTCTTATACAGGTTCATTTGCACCATGG	1140
Qy	381	SerGlyArgPheTrrpSerLeuTrrpAspThrGlyTrrpAlaLysIleHisIleProIleIle	400
Db	1141	AGTGGCAGTTTTATTCATTGTGGGATACCTGGGTATGCAAAATATACATTTCCATATT	1200
Qy	401	AlaSerValSerGluHisGlnProThrTrrpValSerPhePheAspLeuHisIle	420
Db	1201	GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTCTTGTATCTACATTT	1260
Qy	421	LeuValCysThrPheProAlaGlyLeuTrrpPheCysIleLysAsnIleAsnAspGluArg	440
Db	1261	CTTGTATGATCTTCCCGCAGCGGCTTTGGTCTGTCATCAAAATATACAGATGAAGA	1320
Qy	441	ValPheValAlaLeuTrrpAlaIleSerAlaValTrrpPheAlaGlyValMetValArgLeu	460
Db	1321	GTAATTGTGCTCTATATGCAATCAGTGTCTTCTTCTTGTGAGTGTGGTGGTGGTGG	1380
Qy	461	MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe	480
Db	1381	ATGTTGACTTTGACTCCAGTCTGTGTGTGCTGCTGCTGCTGCTTTCCTTCAATGTTTT	1440
Qy	481	GluHisTrrpLeuGlyAspMetLysArgGluAsnProProValGluAspSerSerAsp	500
Db	1441	GAGCATTATTTGGGGGATGACATGAAAGGGGAAATCCACCTGTGGAGGACAGCATGTAT	1500
Qy	501	GluAspAspLysArgAsnGlnGlyAsnLeuTrrpAspLysAlaGlyLysValArgLysHis	520
Db	1501	GAGGATGCAAAAGAAACCAGGAATTTGTATGATAAGCAGGTAAAGTGAAGAAACAT	1560
Qy	521	AlaThrGluGlnGluLysThrGluGlyLeuGlyProAsnIleLysSerIleValThr	540
Db	1561	GCAACTGACAGGAAAACTGAAGAGGATTAGGCCCTTAATAATAAAAGCATTTGTACC	1620
Qy	541	MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrrpValThrSerAsn	560
Db	1621	ATGTTGATGTGATGCTATTGATGATGATGATGATGATGATGATGATGATGATGATGAT	1680
Qy	561	AlaTrrpSerSerProSerValValLeuAlaSerTrrpAsnHisAspGlyThrArgAsnIle	580
Db	1681	GCCTACTAGTCCAGTGTAGTCTTGGCTCATACATCATGATGGCAGGAGATATC	1740
Qy	581	LeuAspAspPheArgGluAlaTrrpPheTrrpLeuArgGlnAsnThrAspGluHisAlaArg	600
Db	1741	TTAGATGATTTTGAAGAGCTTACTTTTGGCTAAGGCAAAATATACAGATGAACATGACCA	1800

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:51:45 ; Search time 9150.76 Seconds
(without alignments)
4268.840 Million cell updates/sec

Title: US-10-028-384-2
Perfect score: 4401
Sequence: 1 MAEPAPESKHKSLSNPSW.....GVKNKLVFKGKISKTKTV 826

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131519_8017/app.query.fasta_1.3740
-DB=genembl -QFMT=fastap -SUFFIX=rge -MINMATCH=0 -LOOCL=0 -LOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USPR=US10028384 @CGN 1.1 20262 @runat_14122004_131519_8017 -NCFU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4401	100.0	2481	6	AX799082 Sequence
2	4401	100.0	2481	9	AX074880 Homo sapi
3	4269.5	97.0	4236	10	BC052433 Mus muscu
4	4263.5	96.9	2710	6	AX799084 Sequence

	5	2950	67.0	2417	3	AF132552
	6	2950	67.0	2417	6	AX799088 Sequence
	7	2950	67.0	2699	6	CQ589353 Sequence
	8	2838	64.5	4922	6	CQ589352 Sequence
	9	2838	64.5	162321	3	AC007853 Drosophil
	10	2838	64.5	181132	3	AC008206 Drosophil
	11	2838	64.5	227219	3	AE003750 Drosophil
	12	2791.5	63.4	75850	2	AC018145 Drosophil
	13	2678	60.8	1734	10	BC003206 Mus muscu
	14	2605	59.2	54118	3	U13019 Caenorhabdi
	15	2370	53.9	1664	6	CQ782554 Sequence
	16	2370	53.9	1664	6	BD127193 Primer fo
	17	2370	53.9	1664	9	AK074587 Homo sapi
	18	2274	51.7	2855	6	CQ573675 Sequence
	19	2266.5	51.5	3094	6	AX799090 Sequence
	20	2266.5	51.5	3094	10	MUSTRPR L34260 Mus musculu
	21	2266	51.5	2736	10	BC037612 Mus muscu
	22	2262	51.4	2118	9	BT007100 Homo sapi
	23	2262	51.4	2118	12	BT008132 Synthetic
	24	2262	51.4	2450	9	BC048348 Homo sapi
	25	2262	51.4	2458	6	CQ834714 Sequence
	26	2262	51.4	2472	6	CQ727867 Sequence
	27	2262	51.4	2516	9	BC020965 Homo sapi
	28	2260	51.4	2293	9	HUMTMC L47337 Homo sapien
	29	2259	51.3	2608	5	BC046072 Danio rer
	30	2258.5	51.3	3686	10	AK129027 Mus muscu
	31	2256	51.3	2481	5	BC067313 Xenopus t
	32	2255	51.2	2898	9	BSM809254 Homo sapi
	33	2253	51.2	2567	5	BC063234 Danio rer
	34	2250	51.1	2472	6	AX799092 Sequence
	35	2250	51.1	2472	9	HUMB5A L38961 Human putat
	36	2240.5	50.9	2435	6	CQ834710 Sequence
	37	2236	50.8	2760	6	CQ42624 Sequence
	38	2188.5	49.7	110000	8	CR382128_21 Continuation (22 o
	39	2171.5	49.3	2555	8	BT012881 Lycopersi
	40	2164	49.2	2256	6	AX489023 Sequence
	41	2147.5	48.8	110000	8	CR382134_03 Continuation (4 of
	42	2138.5	48.4	2756	8	AK099138 Oryza sat
	43	2123	48.2	2819	8	AK067609 Oryza sat
	44	2103.5	47.8	300029	8	AE016903 Eremothec
	45	2096	47.6	110000	8	CR382121_09 Continuation (10 o

ALIGNMENTS

RESULT 1
AX799082
LOCUS AX799082 2481 bp mRNA linear PAT 08-OCT-2003
DEFINITION Sequence 1 from Patent WO03054008.
ACCESSION AX799082
VERSION AX799082.1 GI:37605057
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Perreault, C. and McBride, K.
Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy
AUTHORS Patents: WO 03054008-A 1 03-JUL-2003;
TITLE Compatisgene Inc. (CA)
JOURNAL Location/Qualifiers
FEATURES
source
1..2481
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..2481
/codon_start=1
/product="SIMP"
/protein_id="CAE48447.1"
/db_xref="GI:37605058"
/translation="MAEPAPESKHKSLSNPSWGLMALGNRHHGHGFGAQAHA"

cdna Library Preparation: M. Bento Soares, University of Iowa
 cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: NDAM602 row: i column: 23

High quality sequence stop: 594.

Location/Qualifiers

FEATURES

source

1..709

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:30523894"

/tissue_type="Pooled Chondrosarcoma Tumor cells"

/lab_hosts="DH10B (T1 phage-resistant)"

/clone_lib="NIH_MGC_219"

/note="Vector: pYX-Asc; Site 1: EcoRI; Site 2: NotI;

Library is oligo-dT primed and directionally cloned.

Denatured RNA was size fractionated on a 1% agarose gel.

First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was

size selected according to RNA size fraction, ligated with

EcoR I adaptor, digested with Not I and then cloned

directionally into pYX-Asc vector. Average insert size

0.5-1kb. Adaptors 5' (AATTCGGCAGG)3' and 5'd

(CCTCGTGGCG)3'..3' Linker sequence - GCGGCGCTGAGAGCC T18.

Sequencing primers 3' end: T3 promoter primer 5'd

(ATTAAACCCCTCACTAAGGGA)3'..5' End: T7 promoter primer 5'd

(TATACGACTCTAGG)3'..5' Library was constructed in the

laboratory of M. Bento Soares. Average insert size 2-3kb.

Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.:	4.43e-208	Length:	709
Score:	203.00	Matches:	203
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	24.58%	Indels:	0
DB:	7	Gaps:	0

US-10-028-384-2 (1-826) x CF780700 (1-709)

Qy	608	GlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeuValAspAsnAsnThrTrpAsn	627
Db	9	GGCTATCAGATCGCTGGAATGGCTAATAGAACTACGTTGGTGATTAATAACACCTGGAAT	68
Qy	628	AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLys	647
Db	69	AACAGCCACATAGACCTGGTGGGAAGAAGCTATGCTCTTCTAATGAACACGACCCCTATAAA	128
Qy	648	IleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyr	667
Db	129	ATCATGAGGACTCTAGATGTAGATTTATGTTTGGTTATTTTGGAGGGGTATTGGCTAT	188
Qy	668	SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGlyGlyLysPro	687
Db	189	TCTGGTATGATATCAACAATTTCTCTGGATGGTTAGATAGCTGAGGAGAAATCC	248
Qy	688	LysAspIleArgGlySerAspTyrPheThrProGlnGlyGlyPheArgValAspLysAla	707
Db	249	AAAGACATTCGGAAAGTGAATTTTACCCACAGGAGAAATTCGCTGTAGACAAAGCA	308
Qy	708	GlySerProThrLeuLeuAsnLysLeuMetTyrLysMetSerTyrTyrArgPheGlyGlu	727
Db	309	GGATCCCTACTTGTGTAATTCGCTTATGTATATAAATGTCTATCTACAGATTTGGAGAA	369
Qy	728	MetGlnLeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGly	747
Db	369	ATGAGCTGGATTTTCGTACACCCCGAGGTTTTCACCGAACACGTAATGCTGAGATTGGA	428
Qy	748	AsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuVal	767

Db	429	AATAAGGACATTAAATTTCAACATTGGGAAGGCTTTTACATCAGAACACTGGCTTGT	488
Qy	768	ArgIleTyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgVal	787
Db	489	AGGATATATAAAGTAAAGCACCTGATAACAGGGAGACATTAGATCACAACCTCGAGTC	548
Qy	788	ThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLysArgGly	807
Db	549	ACCAACATTTCCAAAACAGAAATTTTGTCAAGAAGACTACCAAAAAGGAGCGTGGC	608
Qy	808	TyrlleLys	810
Db	609	TACATTAAA	617

Search completed: December 15, 2004, 06:13:58

Job time : 6420.05 secs

RESULT 14
 CF893729
 LOCUS
 DEFINITION
 A0126H08-5 N1A Mouse Undifferentiated ES Cell cDNA Library (Long 1)
 Mus musculus cDNA clone N1A:A0126H08 IMAGE:30726139 5', mRNA
 620 bp mRNA linear EST 04-NOV-2003
 CF893729
 EST.
 CF893729.1 GI:38160778
 Mus musculus (house mouse)
 Mus musculus
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 620)
 Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
 Construction of long-transcript enriched cDNA libraries from
 submicrogram amounts of total RNAs by a universal PCR amplification
 method
 Genome Res. 11 (9), 1553-1558 (2001)
 21429098
 11544199
 Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@gsun.gsc.nia.nih.gov
 Place: A0126 row: H column: 08
 Seq primer: M13 Reverse
 High quality sequence stop: 620
 POLYA=No.
 Location/Qualifiers
 1. 620
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="129/Sv x 129/Sv-CP"
 /db_xref="nia:EST:A0126H08-5"
 /db_xref="taxon:10090"
 /clone="N1A:A0126H08 IMAGE:30726139"
 /dev_stages="R1 ES cells"
 /lab_host="DH10B"
 /clone_lib="N1A Mouse Undifferentiated ES Cell cDNA
 Library (Long 1)"
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
 Site_2: NotI; Mouse cDNA project by the Laboratory of
 Genetics, National Institute on Aging (NIA), Intramural
 Research Program, NIH (http://lgsun.gsc.nia.nih.gov/cDNA).
 This is a long-transcript enriched cDNA library (Ref.
 Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total
 RNAs were obtained from Dr. Kenneth R. Boheler (National
 Institute on Aging, USA). ES cells were cultured without
 feeder cells in the presence of LIF and BRL-conditioned
 media. Double-stranded cDNAs were synthesized with an
 Oligo(dT) primer [Invitrogen]:
 5'-pGACGATGTCAGATCGGACGGCCGCCCTTTT-3' from
 14.2 ug of total RNA, treated with T4 DNA polymerase, and
 purified by ethanol-precipitation. The cDNAs were ligated
 to Lona-linker L1-Sal4, purified by phenol/chloroform, and
 separated from free linkers by Centricon 100. Then, the
 cDNAs were amplified by long-range high fidelity PCR using
 Ex taq polymerase (Takara) with a primer Sal4-S. The
 products were purified by phenol/chloroform and Centricon
 100. The cDNAs were digested with SalI and NotI enzymes
 and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
 vector. The DH10B E. coli host was transformed with the
 ligation mixture by the standard chemical method. The
 average insert size is about 2.4 kb. The library was
 constructed by Yulan Piao."

Alignment Scores:
 Pred. No.: 620
 Score: 206
 Length: 206
 Matches: 206

ORIGIN
 2.2e-211 Length: 620
 206.00 Matches: 206

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.94% Indels: 0
 DB: 7 Gaps: 0
 US-10-028-384-2 (1-826) x CF893729 (1-620)
 QY 260 ValSerAlaIrpGlyGlyTyValPheLeileAsnLeuIleProLeuHisValPheVal 279
 DB 2 GTCTCTGGTGGGAGGTTATGTTCATCATCATCACTCATCCCTCCCATGTGTTGGT 61
 QY 280 LeuLeuMetGlnArgTyrSerLysArgValTyrLeAlaTyrSerThrPheTyrle 299
 DB 62 TTGTGTGTGATGAGAGTACAGAGAGAGTCTACATAGCATATATAGCATTTTCTACAT 121
 QY 300 ValGlyLeuLeuSerMetGlnIleProPheValGlyPheGlnProIleargThrSer 319
 DB 122 GTGGGTTTATATATTCATGAGATACCTTTTGTGGGATTTTCAGCATCATGAGCAAGC 181
 QY 320 GluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
 DB 182 GAGCACATGGCAGCTGCAGGTGCTTTTGGCGTCTGCAAGCTTACGCTTTTTCAGTAT 241
 QY 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
 DB 242 CTGAGACCGGTTGACAAACAGAGAGTTCAGACCCCTTTCTTTTGGGTGTCACATA 301
 QY 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrleAlaPro 379
 DB 302 GCTGACGGCGCTGTGTTCTTCTAGTGTCTATCTGACATACACAGGTTATATTCACCA 361
 QY 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399
 DB 362 TGGAGTGGCAGGTTTATTTCACATGGGATACGCTATGCAAAATATACATCCCAATT 421
 QY 400 IleAlaSerValSerGluHisGlnProThrThrTyrValSerPhePhePheAspLeuHis 419
 DB 422 ATTCATCATGCTGTGAACATGACGCTACGACATGGGTGTCTTCTTCTTTCATCTACAT 481
 QY 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGlu 439
 DB 482 ATTCCTGTGTGTACCTTCCCGAGGCGCTATGCTTCTGCATCAAAATATCAACGATGAA 541
 QY 440 ArgValPheAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
 DB 542 AGAGTGTGTGCTGTGTATGCGATGCGTGTCTACTTTGCCGAGTGATGGTGGCG 601
 QY 460 LeuMetLeuThrLeuThr 465
 DB 602 CTGATGCTGACTCTGACC 619
 RESULT 15
 CF890700
 LOCUS
 DEFINITION
 AGENCOURT 15873364 NIH MGC_219 Homo sapiens cDNA clone
 IMAGE:30523894 5', mRNA sequence.
 ACCESSION
 CF780700
 VERSION
 CF780700.1 GI:37740297
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 709)
 NIH-MGC http://mgs.nci.nih.gov/
 AUTHORS
 National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE
 JOURNAL
 COMMENT
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: James Martin, University of Iowa

```

Db      122 GTAAAGTCTGTGAAGACGGGTCTGTGTTCTGGCAAGATGTGCTGCTGTGCAATATTC 181
Qy      258 TyrMetValSerAlaTfPglyTyrValPheileAseNLeuLeuPheLeuHisVal 277
Db      182 TACATGGTCTCTGGCTGGGAGGTATGTGTTTCATCATCAACCTCATCCCTCTCCATG 241
Qy      278 PheValLeuLeuMetGlnArgTyrSerLysArgValTyrLeuAlaTyrSerThrPhe 297
Db      242 TTGTGTGCTGCTGATGCGAGGTACGCAAGAGTCTACATAGCATATAGCACTTTC 301
Qy      298 TyrLeuValGlyLeuLeuLeuSerMetGlnPheProPheValGlyPheGlnProLeuArg 317
Db      302 TACATTGTGGGTTTAAATATTATCATCATCAGATACCTTTGTGGCAATTCAGCAATCAGA 361
Qy      318 ThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAla-Phele 337
Db      362 ACAAGCGAGCATGGCAGCTGCGAGGTCTTTGGCTGCTGCAAGCTTACCGCTTTTTT 421
Qy      337 uGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyVa 357
Db      422 GCAGTATCTGAGACACGGGTGACAAAACAGGAGTCCAGACCTTTCTTTTGGGTGT 481
Qy      357 lSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrGlyTyrIle 377
Db      482 CTCCTAGCTCAGCGCTGTGTTCTTGTAGTGTCTATCTATCTACATACACAGGTTATAT 541
Qy      377 eAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIle 397
Db      542 TGCACCATGGAGTGGCAGGTTTATTTCATCTATGGGATCTGGGTATGCAAAAATACACAT 601
Qy      397 eProIleLeuAlaSerValSerGluHisGlnProThrTrpValSerPhePheAsn 417
Db      602 TCCAAATATTTCATCAGTGTCTGAAATACATCAGCCCTAGCATGGGTCTCTTCTTTGA 661
Qy      417 pLeuHisIleLeuValCysThr 424
Db      662 TCTACATATCTCTGTATGTACT 683

```

```

RESULT 13
CF728207 637 bp mRNA linear EST 09-OCT-2003
UI-M-HB0-ckm-d-05-0-UI.r1 NIH_BMAP_HB0 Mus musculus cDNA clone
IMAGE:30548716 5', mRNA sequence.
CF728207
CF728207.1 GI:37602375
EST
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 637)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabps-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP).
Seq primer: pYX-5.
Location/Qualifiers
1. .637
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30548716"

```

FEATURES

```

source

```

```

/tissue_type="whole eye"
/dev_stage="embryo 12.5,13.5,14.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_HB0"
/note="Organ: Eye; Vector: pYX-Asc; Site 1: Ecor I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is TTTATTGAGT. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

```

ORIGIN

```

Alignment Scores:
Pred. No.: 7,35e-218 Length: 637
Score: 212.00 Matches: 212
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.67% Indels: 0
DB: 7 Gaps: 0

```

US-10-028-384-2 (1-826) x CF728207 (1-637)

```

Qy      72 LeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAla 91
Db      1 CTCTCTTCCACATCTCTTCTGCTGCTGGTGGCCGCTTCAGCTCGCGCTCTTGGCC 60
Qy      92 ValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgSerThr 111
Db      61 GTCATCCGCTTCGAGAGCATCATCCACGAGTTCGACCGTGGTTAACTATAGATCAACA 120
Qy      112 HisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArgAlaTrp 131
Db      121 CATCATCTTGCATCTCATGGATCTATAGTTCCTAAATTTGGTTGATGAAGAGCATGG 180
Qy      132 TyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrAlaGly 151
Db      181 TACCCACTGGGAAGAATAGTGGTGGCCACCGTTTACCAGCGGTTCATGATAACAGCTGGC 240
Qy      152 LeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspValCysVal 171
Db      241 CTATTTCATTGGATTTTAAATACATTGAACATACAGTTCACATAGAGATGTGTGTGA 300
Qy      172 PheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGlu 191
Db      301 TTCTCTTGCACCAACTTTTAGCGGCTTACATCCATATCTAGCTTCCTGCTAACTAGAGAA 360
Qy      192 LeuTrpAsnGlnGlyValGlyLeuLeuAlaCysPheIleAlaIleValProGlyTyr 211
Db      361 CTGTGAACCAAGAGCAGGAGCTTCTAGTCTGCTGCTTCATTTGCTATCTGACAGGATAC 420
Qy      212 IleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGln 231
Db      421 ATATCTCGGTGATGCGGGGATCTCTTGTATATGAAGGCAATTCGCAATTTTTCGCTTCAG 480
Qy      232 PheThrTyrTyrLeuTrpValLysSerValLysThrGlySerValPheTrpThrMetCys 251
Db      481 TTCATCTTACTATTATGGGTAAAGTCTGTGAAGACCGGGTCTGTGTTCTGACACATGTGC 540
Qy      252 CysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsn 271
Db      541 TGCTGCTTGTTCATATTTCTACATGTGCTCTCGTGGGAGGTTATGTGTTTCATCATCAAC 600
Qy      272 LeuIleProLeuHisValPheValLeuLeuLeuMet 283
Db      601 CTCAATCCCTCTCCATGTGTGTTGTGTTGCTGTGATG 636

```


feeder cells in the presence of LIF and BRL-conditioned media. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen].
 5'-pGACTAGTTCTAGATCGGAGCGCGCCCTTTTTTTTTTTT-3' from 14.2 ug of total RNA, treated with Ta DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lons-linker Lu-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.4 kb. The library was constructed by Yulan Piao."

ORIGIN

Alignment Scores:
 Pred. No.: 5,11e-220 Length: 645
 Score: 214.00 Matches: 214
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 25.91% Indels: 0
 DB: 7 Gaps: 0

US-10-028-384-2. (1-826) x CF893672 (1-645)

QY 260 ValSerAlaTrpGlyGlyTyrValPheIleAseLeuLeuProLeuHisValPheVal 279
 Db 2 GTCCTGCGTGGGAGGTATGTTTCATCATCACTCATCTCCCTCCATGTTGTTGG 61
 QY 280 LeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIle 299
 Db 62 TTGCTGCTGATGCGAGGTACAGACAGAGTCTTACATAGCATATAGCACTTCTACATT 121
 QY 300 ValGlyLeuLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
 Db 122 GTGGGTTTAAATATTATTCATCATCATCTTTTGTGGATTTTCAGCAATCATCAAGC 181
 QY 320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
 Db 182 GAGCACATGGGAGCTGCGAGTGTCTTTGCGCTGCTGCAAGCTTACCTTTTGTGCAAT 241
 QY 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
 Db 242 CTGAGAGACCGGTTGACAAACAGAGTTCAGACCCCTTTCTTTTGGGTGTCACATA 301
 QY 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
 Db 302 GCTGAGGCGGTGTGTTCTTCTTAGTGTCTATCTATGACATACACAGGTATATGACCA 361
 QY 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399
 Db 362 TGGAGTGCAGGTTTATTCTACTATGGATCTGGTATCGGATCGAAATATACATCCATT 421
 QY 400 IleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePhePheLeuHis 419
 Db 422 ATTGCATCAGTGTCTGAACATCACCCCTACGACATGGGTGCTTCTTCTTCTTGTATCAT 481
 QY 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGlu 439
 Db 482 ATCTCTGTATGATCTTCCAGCAGGCTATGGTCTGATCAAAAATATCAACGATGA 541
 QY 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
 Db 542 AGAGTGTTTGGTCTGTATGCGCATCAGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 601
 QY 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAla 473
 Db 602 CTGATGCTGACTCTCACCCCGGTCTGCTGATGCTGTGCGCC 643

RESULT 12

CF742714
 LOCUS
 DEFINITION UI-M-HBO-clm-m-17-0-UI.r1 NIH_BMAP_HBO Mus musculus cDNA clone
 IMAGE:30618064 5', mRNA sequence.
 CF742714
 ACCESSION
 VERSION CF742714.1 GI:37639053
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 692)
 REFERENCE
 1 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 UNPUBLISHED (1999)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-femail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
 http://genome.uiowa.edu/distribution/mousefl.html
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pyX-5.

Location/Qualifiers
 1. 692
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:30618064"
 /tissue_type="whole eye"
 /dev_stage="embryo 12.5,13.5,14.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP_HBO"
 /notes="Organ: Eye; Vector: pyX-Asc; Site 1: Ecor I;
 Site 2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with Ecor I adaptor, digested with NotI and then cloned
 directionally into pyX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is TTTATGAGCT. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the
 Developing Mouse Nervous System", supported by National
 Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:
 Pred. No.: 5,5e-220 Length: 692
 Score: 214.00 Matches: 227
 Percent Similarity: 99.56% Conservatives: 0
 Best Local Similarity: 99.56% Mismatches: 0
 Query Match: 25.91% Indels: 1
 DB: 7 Gaps: 0

US-10-028-384-2 (1-826) x CF742714 (1-692)

QY 198 GlyLeuLeuAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValala 217
 Db 2 GGATCTTCTAGCTGCTTCTTCTATGCTATCGTACCGGATCATATCTCGTCAGTGGCG 61
 QY 218 GlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrp 237
 Db 62 GGATCCCTTGTATTAATGAAGCATTCATTTTGGCTTCAGTTCATCTACTACTATG 121
 QY 238 ValLysSerValLysThrGlySerValPheThrMetCysCysLeuSerTyrPhe 257

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 National Cancer Genomics
 Office of Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Stefan Hanson
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
 and advice from Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDAM615 row: k column: 08
 High quality sequence stop: 625.
 FEATURES Location/Qualifiers
 1..861
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30528919"
 /tissue_type="Human Placenta"
 /lab_hosts="DH10B TorA"
 /clone_lib="NIH_MGC 147"
 /note="Organ: placenta; Vector: pBluescriptR; Site: 1:
 all-XhoI; Site 2: BamH; Oligo-dT primed using primer
 5'-TTTTTTTTTTTTTNN-3', size-selected for average
 insert size 2.3 kb and normalized to 500. This is a
 primary library enriched for full-length clones and
 constructed using the Cap-trapper method (Carninci, in
 preparation). Library constructed by M. Brownstein
 (NIH/NHGRI, National Institutes of Health). Note: This is
 a NIH_MGC library."
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,69e-225 Length: 861
 Score: 219.00 Matches: 232
 Percent Similarity: 99.57% Conservative: 0
 Best Local Similarity: 99.57% Mismatches: 0
 Query Match: 26.51% Indels: 1
 DB: 7 Gaps: 0
 US-10-028-384-2 (1-826) x CF995035 (1-861)
 QY 561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
 Db 27 GCCTACTCTAGTCCAAAGTGTAGTCCCTGCTCATAAATCATGATGGCCACGAGGAATATC 86
 QY 581 LeuAspAspPheArgGluAlaTyrPheTyrLeuArgGlnAsnThrAspGluHisAlaArg 600
 Db 87 TTAGATGATTTTACAGAGCTTACTTTTGCTTAAGCAAAATACAGATCAATGACGCA 146
 QY 601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
 Db 147 GTAATGCTCTGGTGGGATTATGGCTATCATAGATAGCTGGAATGGCTAATAGAACTACGTG 206
 QY 621 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
 Db 207 GTGGATAATACACCTCGAATAACAGCCACATAGCACTGGTGGGAAAGCTATGCTTCT 266
 QY 641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValleuValle 660
 Db 267 AATGAAACAGCAGCCTATAAAATCATGAGACCTCTAGATGTAGATTATGTTGTTATT 326
 QY 661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
 Db 327 TTTCGAGGGGTATTGCTGCTATCTGGGTGATGATATCAACAAATTTCTCTGGATGTAGG 386
 QY 681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700

Db 387 ATAGCTGAAGGAGAACATCCCAAGACATTCGGAAAGTACTATTTTACCCCAAGGGA 446
 QY 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrIleMet 720
 Db 447 GAATTCGGTGTAGACAAAGCAGGATCCCTACTTTGTTGAATTCGCTTATGTATATAAATG 506
 QY 721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAsnArg 740
 Db 507 TCATCTACAGATTTCGAGAATTCAGCTGGGATTTTCGTACACCCCAAGGTTTTCACCGGA 566
 QY 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPhe 760
 Db 567 ACACGTAATGCTGAGATTGGAATAAGGACATTAATAATCAACATTTTGAAGAGCGCTTT 626
 QY 761 ThrSerGluHisTrpLeu-ValArgIleTyrLysValLysAlaProAspAsnArgGluTh 780
 Db 627 ACATCAGAACACCTGGCTTTGTTAGGATATATAAAGTAAAGCACCTGATAACAGGAGAC 686
 QY 780 rLeuAspHisLysProArgValThrAsnIlePhePro 792
 Db 687 ATTAGATCAAAACCTCGAGTCACCAACATTTTCCC 723
 RESULT 11
 CF893672
 LOCUS 645 bp mRNA linear EST 04-NOV-2003
 DEFINITION A0126C03-5 NIA Mouse Undifferentiated ES Cell cDNA Library (Long 1)
 Mus musculus cDNA clone NIA:A0126C03 IMAGE:30726074 5', mRNA
 sequence.
 ACCSSION CF893672
 VERSION CF893672.1 GI:38160721
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
 1 (bases 1 to 645)
 Construction of long-transcript enriched cDNA libraries from
 submicrogram amounts of total RNAs by a universal PCR amplification
 method
 Genome Res. 11 (9), 1553-1558 (2001)
 MEDLINE 21429098
 PUBMED 11544199
 COMMENT Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdnas@igsun.grc.nia.nih.gov
 Plate: A0126 row: C column: 03
 Seq primer: M13 Reverse
 High quality sequence stop: 645
 POLYA=No.
 FEATURES Location/Qualifiers
 1..645
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="129/Sv x 129/Sv-Cp"
 /db_xref="niaEST:A0126C03-5"
 /db_xref="taxon:10090"
 /clone="NIA:A0126C03 IMAGE:30726074"
 /dev_stage="R1 ES cells"
 /lab_host="DH10B"
 /clone_lib="NIA Mouse Undifferentiated ES Cell cDNA
 Library (Long 1)"
 /note="Vector: PCMV-SpORT6 (Invitrogen); Site 1: SalI;
 Site 2: NotI; Mouse cDNA project by the Laboratory of
 Genetics, National Institute on Aging (NIA), Intramural
 Research Program, NIH (<http://igsun.grc.nia.nih.gov/cDNA>).
 This is a long-transcript enriched cDNA library (Ref.
 Genome Res. 11, 1553-1558 (2001). (PMID: 11544199)). Total
 RNAs were obtained from Dr. Kenneth R. Boheler (National
 Institute on Aging, USA). ES cells were cultured without

us-10-028-384-2.Oligo.rst

Thu Dec 16 16:25:06 2004

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Dr. David Rowe
 cDNA Library Preparation: Invitrogen Corp
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 DNA Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Alignment Scores: 2,748-237 Length: 693
 Pred. No.: 230.00 Matches: 230
 Score: 100.00% Conservativity: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 27.85% Indels: 0
 Query Match: 7 Gaps: 0
 DB:

US-10-028-384-2 (1-826) x CN461211 (1-693)

QY 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 DB 3 ACAGTTTCACATAAGAGATGTGTGTATTCCTTGCACCAACTTTTAGCGGCTTACATCC 62
 QY 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 202
 DB 63 ATATCTACGTTCTCTGTAACAGAACTGTGGAAACCAAGGAGGAGGACTTCTAGCTGCT 122
 QY 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn 222
 DB 123 TGTCTTCATTGCTATGTCACAGGTACATATCTCGTCACTGCGGAGTCTTTCATTAAT 182
 QY 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrLeuTrpValValSerVallys 242
 DB 183 GAAGGCAATTCGCCATTTTTCGCTTCAGTTCACTTACTTATGGGTAAAGTCTGTGAAG 242
 QY 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
 DB 243 ACCGGTCTGTGTTCTGGACATGTCTGCTGTTCATATTTCTACATGCTCTGCG 302
 QY 263 TrpGlyGlyTyrValPheIleLeuLeuLeuProLeuHisValPheValLeuLeuLeu 282
 DB 303 TGGGAGGTTATGTGTTCATCACTCAACCTCATCCCTCTCCATGTGTGTGTGTGTG 362
 QY 283 MetGluArgTyrSerIleValGlyValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
 DB 363 ATGACAGGTACAGACAGAGAGTCTACATAGCATATGACATTTCTACATCTGGGT 422
 QY 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 DB 423 ATATTATCCATGCAGATACCTTTTGTGGGATTTTCAGCCAAATCAGAACAGCAGC 482
 QY 323 AlaAlaIleValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
 DB 483 GCAGCTCGAGGTGTCTTTCGCTGCTGCAAGCTTACGCTTTTTCGAGTATCTGAGAC 542
 QY 343 ArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaIleGly 362
 DB 543 CGGTTGACAAAACAGGAGTTCAGACCCCTTTCTTTTGGGTCTCTACATGCTGAGGC 602
 QY 363 AlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProThrPheSerGly 382
 DB 603 GCTGTGTTCTCTTAGTGTCTATCTGACATACAGGTTATATGACCATGGAGTGGC 662
 QY 383 ArgPheTyrSerLeuTrpAspThrGlyTyr 392
 DB 663 AGGTTTATTCACATGCGGATCTGGGTAT 692

RESULT 8
 BU515954
 LOCUS AGNCCOURT 10120433 NIH MGC 134 Mus musculus cDNA clone
 DEFINITION IMAGE:6512642 5', mRNA sequence.
 BU515954
 ACCESSION BU515954.1 GI:22823480
 VERSION EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 950)
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/.

TITLE
 JOURNAL
 COMMENT

FEATURES
 source

1..950
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6512642"
 /tissue_type="undifferentiated limb"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_134"
 /notes="Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
 Cloned unidirectionally. Primer: Oligo dT. Average insert
 size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
 this is a NIH_MGC Library."

ORIGIN

Alignment Scores: 6,648-234 Length: 950
 Pred. No.: 227.00 Matches: 240
 Score: 99.59% Conservativity: 0
 Percent Similarity: 99.59% Mismatches: 0
 Best Local Similarity: 27.48% Indels: 1
 Query Match: 5 Gaps: 0
 DB:

US-10-028-384-2 (1-826) x BU515954 (1-950)

QY 215 SerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 234
 DB 1 TCAGTGGCGGATCCCTTGTATATGAAGGCAATGCCATTTTCGGCTTCAGTTCATTAC 60
 QY 235 TyrLeuTrpValLysSerValLysThrGlySerValPheTrpThrMetCysCysLeu 254
 DB 61 TACTTATGGGTAAAGTCTGTGAAGACCGGCTCTGTCTCGACAATGTCTGCTGCTTG 120
 QY 255 SerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleLeuLeuLeuPro 274
 DB 121 TCATATTTCATATGTTCTCTGGTGGGAGGTTATGTGTTTCATCATCAACCTCATCCCT 180
 QY 275 LeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyr 294
 DB 181 CTCCATGTGTTTGTGTGCTGCTGATGCAGAGGTACAGCAAGAGAGTCTACATAGCATAT 240
 QY 295 SerThrPheTyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPheGln 314
 DB 241 AGCACTTTCTACATGTGGGTTTAATATATCCATGCAATACCTTTTGTGGGATTTTTCAG 300
 QY 315 ProIleArgThrSerGluHisMetAlaAlaIleGlyValPheAlaLeuGlnAlaTyr 334
 DB 301 CCATCATCAACAAGGAGCAGCATGCGAGTGTCTTGGCTGTCTGCAAGCTTAC 360
 QY 335 Ala-PheLeuGlnTyrLeuArgAspArgLeuThrIleGlnGluPheGlnThrLeuPhePh 354
 DB 361 GCTTTTTTTCAGATATCTGAGAGACCGGTGACAAAAACAGAGGTTCCAGACCCCTTTCTT 420
 QY 354 eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyr 374
 DB 421 TTTGGGTCTCTCACTAGCTGAGCGGCTGTGTTCTCTTAGTGTCTATCTATCTGCATACAC 480
 QY 374 cGlyTyrIleAlaProThrPheSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLy 394
 DB 481 AGGTATTATTCACCATGAGTGGCAGGTTTATTATTCTACTATGGGATATCTGGGTATGCAA 540

Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaaps-remail.nih.gov
Tissue Procurement: Naryan Bhat
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NEAM1098 row: d column: 22
High quality sequence stop: 710.

FEATURES

Location/Qualifiers
1..775
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30786813"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_222"
/notes="Organ: placenta; Vector: pExpress-1; Site 1: EcoRV;
Site 2: NotI; RNA obtained from three placentas from
female C57/BL6 mouse at 16 days pregnancy. Tissues were
snap-frozen and kept at -80C for two days before RNA
extraction and purification (Tri-reagent method). cDNA was
primed using oligo-dT primer:
5'-TGACTAGTTCTAGATCGGAGCGCGCCG(T)25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. Size-selection >1 kb
resulted in an average insert size of 1.5 kb. Library is
not amplified. (Normalized version of this library is
NIH MGC 203.) Library constructed by Express Genomics
(Frederick, MD)."

ORIGIN

Alignment Scores:
Pred. No.: 2,11e-239 Length: 775
Score: 232.00 Matches: 232
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 28.09% Indels: 0
Gaps: 0
DBs: 7

US-10-028-384-2 (1-826) x CK032376 (1-775)

Qy 194 AsnGlnGlyAlaGlyLeuLeuAlaAaCysPheIleAlaIleValProGlyTyrIleSer 213
Db 2 AACCAAGGAGGAGGACTTCTAGCTGCTTCTTCTATGCTATGACGAGGATACATATCT 61
Qy 214 ArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 233
Db 62 CGGTCAAGTGGCGGATCTTTGATATGAAGGAGGATGCGCATTTTGGCTTCAGTTCACT 121
Qy 234 TyrTyrLeuTrpValLysSerValLysThrGlySerValPheThrMetCysCysCys 253
Db 122 TACTACTTATGGTAAAGTCTGTGAAGCGGGTCTGTGTCTGGCAATGTGCTCTGTC 181
Qy 254 LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIle 273
Db 182 TTGTCAATTTCTACATGCTCTCGGTGGGAGGTTATGTTCTATCATCACTCATC 241
Qy 274 ProLeuHisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrIleAla 293
Db 242 CCTCTCCATGTGTTTGTGCTGCTGATGACAGAGGTACAGAGAGGCTTACATAGCA 301
Qy 294 TyrSerThrPheTyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPhe 313
Db 302 TATACACTTTCTACATTTGGGTTTATATATCAATGAGATACCTTTTGGGATTT 361
Qy 314 GlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAla 333
Db 362 CAGCCAAATCAGAAAGGAGGACATGCGAGCTGAGGTGCTTTGGCTGTGCTGAAGCT 421

Qy 334 TyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhe 353
Db 422 TAGGCTTTTTCAGTATCTGAGAGCCGGTTGACAAACAGGAGTTCCACACCTTTTC 481
Qy 354 PheLeuGlyValSerLeuAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyr 373
Db 482 TTTTGGGTTCTCACTAGCTGACGCGCTGTGTTCCCTTAGTGTCACTATCTGACATAC 541
Qy 374 ThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 393
Db 542 ACAGGTATATGTCACCATGGAGTGGCGGTTTATTTCATGATGGATACCTGGGTATGCA 601
Qy 394 LysIleHisIleProIleAlaSerValSerGluHisGlnProThrThrTyrValSer 413
Db 602 AAAATACATTCCTCAATATTGATCAGTGTCTGACATCAGCTTACGACATGGGTGCT 661
Qy 414 PhePhePheAspLeuHisIleLeuValCysThrPhe 425
Db 662 TTCTTCTTGTATCTACATATCTTGTATGATACCTTC 697

RESULT 7

CN461211 693 bp mRNA linear EST 21-APR-2004
UI-M-HNO-cot-g-05-0-UI.r1 NIH_BMAP_HNO Mus musculus cDNA clone
IMAGE:30652468 5', mRNA sequence.
CN461211
CN461211.1 GI:46466937
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 693)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP).

Seq primer: pYX-5.

FEATURES

Location/Qualifiers
1..693
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30652468"
/tissue_type="Upper Head"
/dev_stage="9.5-10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_HNO"
/notes="Organ: Head; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CGAATGAT. This library was created for the University
of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

ORIGIN

QY	374	ThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla	393
Db	543	ACAGGTTATTGACCATGGAGTGGCAGGTTTATTCTACATGGGATCTGGGTATGCA	602
QY	394	LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpValSer	413
Db	603	AAATACACATTCCTCAATATTGATCAGTGCTGAACTACGCTTACGACATGGGTGCT	662
QY	414	PhePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIle	433
Db	663	TTCTTCTTTGACTACATATTCTTGATGTACTCTCCAGCAGGCTATGGTTCTGCATC	722
RESULT 5			
BE952197		747 bp mRNA linear EST 29-APR-2002	
LOCUS		UI-M-CDO-aj1-h-06-0-UI.61 NIH BMAP Ret2 Mus musculus cDNA clone	
DEFINITION		UI-M-CDO-aj1-h-06-0-UI 3', mRNA sequence.	
ACCESSION		BE952197	
VERSION		BE952197.1 GI:10592357	
KEYWORDS		EST.	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
REFERENCE			
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
JOURNAL		Bonaldi M.F., Lemmon G. and Soares M.B.	
MEDLINE		1 (bases 1 to 747)	
PUBMED		Normalisation and subtraction: two approaches to facilitate gene	
COMMENT		discovery	
		Genome Res. 6 (9), 791-806 (1996)	
		97044477	
		889548	
		Contact: Chin, H	
		National Institute of Mental Health	
		6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD	
		20892-9643, USA	
		Tel: 301 443 1706	
		Fax: 301 443 9890	
		Email: MEST@mail.nih.gov	
		Oligo-dT track not found. Not I site shown in beginning of sequence	
		is likely internal to the message. cDNA Library Preparation: M.B.	
		Soares Lab Clone distribution: Researchers may obtain BMAP cDNA	
		clones from RESEARCH GENETICS. It should be noted that Bento Soares	
		is generating a small number of additional specialized	
		non-redundant arrays of BMAP cDNAs whose availability will be	
		considered under appropriate and limited collaborative arrangements	
		The tissue for this library was contributed by Dr. Xin-Yuan Fu,	
		Yale University School of Medicine	
		Seq primer: M13 Forward	
FEATURES			
source		Location/Qualifiers	
		1..747	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="C57BL/6J"	
		/db_xref="taxon:10090"	
		/clone="UI-M-CDO-aj1-h-06-0-UI"	
		/dev_stage="1 day"	
		/lab_host="DH10B (Life Technologies)"	
		/clone_lib="NIH BMAP Ret2"	
		/note="Vector: pT7m3D-Pac (Pharmacia) with a modified	
		polylinker; Site 1: Not I, Site 2: Eco RI; The	
		NIH BMAP Ret2 library is derived from mouse neonatal	
		retina tissue. For a detailed description of the library	
		from which this clone was derived, please visit our web	
		site at brainest.eng.uiowa.edu. The tissue for this	
		library was contributed by Dr. Xin-Yuan Fu, Yale	
		University School of Medicine	
		TAG_SEQ=None found"	
ORIGIN			
Alignment Scores:			
Pred. No.:		1.16e-242	Length: 747

Score:	235.00	Matches:	235
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	28.45%	Indels:	0
DB:	2	Gaps:	0
US-10-028-384-2 (1-826) x BE952197 (1-747)			
QY	157	LeuAsnThrLeuAsnIleThrValHisIleAraCAspValCysValPheLeuAlaProThr	176
Db	24	TTAAATACATTGAACATAACAGTTACATAAGAGATGTGTGTATTCCTTGACCAACT	83
QY	177	PheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGly	196
Db	84	TTTAGCGGCTTACATCCATATCTACTGTTCTTCTTAAGTAAGTGTGGAACTGTGGAA	143
QY	197	AlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerVal	216
Db	144	GCAGGACTTCTAGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	203
QY	217	AlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeu	236
Db	204	GCGGATCCCTTGAATATGAGCATGTCATTTTGGCTTCAGTTCTACTTACTTACTTA	263
QY	237	TrpValYssSerValYssThrGlySerValPheTrpThrMetCysCysCysSerTyr	256
Db	264	TGGTAAAGTCTGTGAAGACCGGCTGTGTCTTCTGACAAATGTGTCTGCTTGTCTAT	323
QY	257	PheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHis	276
Db	324	TTCTACATGCTCTGCGTGGGAGGTTATGTTTCTATCATCAACCTCATCCTCTCCAT	383
QY	277	ValPheValLeuLeuLeuMetGlnAArgTyrSerLysArgValTyrIleAlaTyrSerThr	296
Db	384	GTGTTTGT	443
QY	297	PheTyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIle	316
Db	444	TTCTACATTTGGGTTTAAATATTCATCCATGACATACCTTTTGGGATTTAGCCAAATC	503
QY	317	ArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPhe	336
Db	504	AGAACAGCCAGCAGACATGGCAGTGCAGGTGTCTTTGGCTGTGCAAGCTTACGCTTT	563
QY	337	LeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGly	356
Db	564	TTGCAGTATCTGAGAGACCGGTTTGACAAAACAGAGTTCAGACCCCTTTTCTTTTGGGT	623
QY	357	ValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyr	376
Db	624	GTCTCACTAGTGCAGCGGCTGTGTTCTTCTTGTGTCTATCTATCTGACATACACAGTTAT	693
QY	377	IleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGly	391
Db	684	ATTGCACATGAGTGGCAGGTTTATTCTACTATGGGATCTGGG	728

CK032376 775 bp mRNA linear EST 26-NOV-2003
AGENCOURT 16650470 NIH MGC_222 Mus musculus cDNA clone
IMAGE:30786813 5', mRNA sequence.
CK032376
CK032376.1 GI:38558300
EST.
Mus musculus (house mouse)
Mus musculus
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 775)
NIH-MGC <http://mgs.nci.nih.gov/>
National Institutes of Health
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.

```

Db 1683 CACTGCACGTGGGTCAAGCAACGCTACTCCAGTCCAAAGTGTGGCTCTGCTCTAC 1742
Qy 573 AsnHisAspGlyThrArgAsnHisLeuAspAspPheArgGluAlaTyrPheTyrLeuArg 592
Db 1743 AATCATGATGTTACCGAATATATTAGATGATTTAGAGAAGCGTACTTTGGCTGAGA 1802
Qy 593 GlnAsnThrAspGluHisAlaArgValMetSerTyrPheTyrPheTyrGlyTyrGlnHisLeuAla 612
Db 1803 CAATAACACGATGATGAACACGCGCGGTTCATGCTGGTGGGACTACGCGTATCATGATTGCT 1862
Qy 613 GlyMetAlaAsnArgThrThrLeuValAspAsnThrTyrPheAsnHisLeuAla 632
Db 1863 GGATGCGCCACAGGACCATCTGTTGGATGAACAACCTGGAAACACACCCATCGCA 1922
Qy 633 LeuValGlyAlaMetSerSerAsnGluThrAlaAlaTyrLysLeuMetArgThr-Le 652
Db 1923 CTGTCGGAAGAAGCATGTCTTCCAATGAACGCGCGCTATATAAATCATGAGTC-CCT 1981
Qy 652 uAspValAspTyrValLeuValLlePheGlyGlyValLleGlyTyrSerGlyAspAsp 1981
Db 1982 TGATGTCGATTAATGTTGTTGTTATTTCCGAGGAGTATGGCTATTCGGGACGATAT 2041
Qy 672 eAsnLysPheLeuTyrMetValArgLleAlaGluGlyGluHisProLysAspLleArg 692
Db 2042 CAACAAGTTCCTCTGGATGTCAGGATAGCTGAAGGGGAGCATCCCAAGACATCCGGA 2101
Qy 692 uSer-AspTyrPheThrPro-GinglyGluPheArgValAspLysAlaGlySerProThr 711
Db 2102 AGG-TGACTATTTCAC-CCAGCAGGAGAGTCCGAGTAGACAAGCTGGGTCTCCTACT 2159
Qy 712 LeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAsp 731
Db 2160 CTGTAAACGCTTATGATATAAATGTCATACATCAGATTGGAGAAATGACGATAGT 2219
Qy 732 PheArgThrProProGlyPheAspArgThrArgAsnAlaGluLleGlyAsnLysAsp 751
Db 2220 TTTCGCACCTCCCGCGCTTGACCAACACGTAATGCTGAGATGGAATAAAGACATT 2279
Qy 752 LysPheLysHisLeuGluAlaPheThrSerGluHisTyrPheValArgLleTyrLys 771
Db 2280 AAATTCAAGCATTTGAGGAGAGCTTTTACATCAGACCATGCTGTCAGGATATATAA 2339
Qy 772 ValLysAlaProAspAsnArgGluThrLeu 781
Db 2340 GTGAAGACCTGACAAACAGGGAGACACTA 2369

```

```

RESULT 4
CN458732
LOCUS
DEFINITION
  CN458732 852 bp mRNA linear EST 21-APR-2004
  UI-M-HBO-cov-1-02-0-UI-r1 NIH_EMAP_HBO Mus musculus cDNA clone
  IMAGE:30649513 5', mRNA sequence.
  CN458732
  EST.
  CN458732.1 GI:46464458
  Mus musculus (house mouse)
  Mus musculus
  ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 bases 1 to 852
  NIH-MGC http://mgi.nci.nih.gov/
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgabers-r@mail.nih.gov
  Tissue Procurement: Dr. James Lin University of Iowa
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Distribution information can be found at
  http://genome.uiowa.edu/distribution/mousefl.html
  This clone was contributed by the Brain Molecular Anatomy Project
  (BMAP)

```

FEATURES

```

Seq primer: pYX-5.
Location/Qualifiers
  1..852
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="CS7BL/6"
  /db_xref="taxon:10090"
  /clone="IMAGE:30649513"
  /issue_type="whole eye"
  /dev_stage="embryo 12.5, 13.5, 14.5 dpc"
  /lab_host="DH10B (TI phage resistant)"
  /clone_lib="NIH_EMAP_HBO"
  /note="Organ: Eye; Vector: pYX-Asc; Site 1: Ecor I;
  Site 2: Not I; The library was constructed according
  Bernaldo, Lennon and Soares, Genome Research, 6:791-806,
  1996. Denatured RNA was size fractionated on a 1% agarose
  gel. First strand cDNA synthesis was primed with oligo-dT
  primer containing a Not I site. Double strand cDNA was
  size selected according to mRNA size fraction, ligated
  with Ecor I adaptor, digested with Not I and then cloned
  directionally into pYX-Asc vector. The library tag
  is TTAATTGAAGT. This library was created for the University
  Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
  Developing Mouse Nervous System', supported by National
  Institute of Mental Health (NIMH)."
```

ORIGIN

```

Alignment Scores:
Pred. No.: 5,19e-248 Length: 852
Score: 240.00 Matches: 240
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.06% Indels: 0
DB: 7 Gaps: 0
US-10-028-384-2 (1-826) x CN458732 (1-852)

```

```

Qy 194 AsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheAlaLeuValProGlyTyrIleSer 213
Db 3 AACCAAGAGCAGGACCTCTAGCTGCTGCTTCAATGCTATCTGACAGGATCATATCT 62
Qy 214 ArgSerValAlaGlySerPheAspAsnGluGlyLeuAlaAlaPheAlaLeuGlnPheThr 233
Db 63 CGGTCACTGGCGGATCTTTGATATGAAGAGCATTTGCCATTTTGGCTTCACTTCACT 122
Qy 234 TyrTyrLeuTrpValLysSerValLysThrGlySerValPheTyrPheTyrMetCysCys 253
Db 123 TACTACTATATGGGTAAAGTCTGTGAAGACCGGCTCTGTGTTCTGGACAATGTCTGCTGC 182
Qy 254 LeuSerTyrPheTyrMetValSerAlaTyrGlyTyrValPheIleIleAsnLeuLeu 273
Db 183 TTGTCATATTTTACATGGTCTCTGGTGGGAGGTATGTGTTCATCATCAACCTCATC 242
Qy 274 ProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAla 293
Db 243 CCTCTCCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 302
Qy 294 TyrSerThrPheTyrIleValGlyLeuLeuLeuSerMetGlnIleProPheValGlyPhe 313
Db 303 TATAGACATTTTCTACATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 362
Qy 314 GlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuGlnAla 333
Db 363 CAGCAATCAGAAACAACGAGGACATGGCAGCTGCGAGGTGCTTTTGGCTGCTGCAAGCT 422
Qy 334 TyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhe 353
Db 423 TAGCTTTTTCAGTATCTGAGAGACCGGTGACAAAACAGGAGTTCCAGACCCCTTTTC 482
Qy 354 PheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyr 373
Db 483 TTTTGGGTGCTCACTAGCTGCGCGCTGTTTCTTCTAGTGTCTATCTATCTGACATAC 542

```


ORIGIN

Alignment Scores: 5,24e-293 Length: 2212
 Pred. No.: 282.00 Matches: 282
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 34.14% Gaps: 0
 DB: 3

US-10-028-384-2 (1-826) x CR610915 (1-2212)

Qy	545	MetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSer	564
Db	3	ATGCTATTGATGATGTTTGGTCTGCTACCTGGGTCAAGCAATGGCTACTTAGT	62
Qy	565	ProSerValLeuAlaSerTyrAsnHisAspGlyThrArgAsnLeuLeuAspAspPhe	584
Db	63	CCAAAGTGTAGTCTGGCTGCCTCATCAATCATGATGGCCACCAAGGAATATCTTAGATGATTTT	122
Qy	585	ArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrp	604
Db	123	AGAGAGGCTTACTTTGGCTAGGCAAAATACAGATGACATGCACGAGTAATGCTTTGG	182
Qy	605	TrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrTrpLeuValAspAsnAsn	624
Db	183	TGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGGTGGATAATAAC	242
Qy	625	ThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAla	644
Db	243	ACCTGGATACAGACCATATAGCTGCTGGGAAAGCACTATGCTTCTAATGAAACAGCA	302
Qy	645	AlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyVal	664
Db	303	GCCTATAAAATCATGAGGACTCTAGATGATGATTTGTTGGTATTTTGGAGGGGTT	362
Qy	665	IleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGly	684
Db	363	ATGGCTATTCTGGTGAATGATACAAATTTCTTGGATGGTTAGTAGTGAAGCA	422
Qy	685	GluHisProLysAspIleArgLysSerAspTyrPheThrProGlnGlyLysPheArgVal	704
Db	423	GAACTCCCAAGACATTCGGGAAAGTCACTATTATCCCAAGAGGAGAAATCCCTGTA	482
Qy	705	AspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTrpArg	724
Db	483	GACAAAGCAGGATCCCTACTTTGTTGATTCCTTATGATATAAATGTCATCTACAGA	542
Qy	725	PheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsnAla	744
Db	543	TTTGAGAAATGCAGCTGGATTTTGTACACCCCGAGGTTTTCACCGAACACGTAATGCT	602
Qy	745	GluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPheThrSerGluHis	764
Db	603	GGATGGAAATAGCAATTAATTCAAATTCATTTGGAAGAGCCCTTATCATCAGAACAC	662
Qy	765	TrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLys	784
Db	663	TGGCTTTTAGGATATATAAAGTAAAGCACCTGATACACGGGAGACATTAGATCACAAA	722
Qy	785	ProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArg	804
Db	723	CTTCAGTACCAACATTTTCCCAACAGAGATTTTGTCAAGAGAGACTACCAAAAGG	782
Qy	805	LysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLys	824
Db	783	AAGCGTGGCTACATATAAATAAGTGGTTTTTAAAGAAAGCAAGAAATATCTAAGAG	842
Qy	825	ThrVal	826
Db	843	ACTGTT	848

RESULT 3

AK012153

LOCUS

DEFINITION

AK012153 2669 bp mRNA linear HTC 03-APR-2004
 Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
 enriched library, clone:2610524N02 product:hypothetical
 Oligosaccharyl transferase (OTase) STT3 subunit containing protein,
 full insert sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AK012153.1 GI:12848721

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

PUBMED

10349636

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

COMMENT

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 2669)

Adachi, J., Aizawa, K., Akahita, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasekawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-vesegsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

Please visit our web site (<http://genome.gsc.riken.jp/>) for further

QY 255 SerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleLeuLeuIlePro 274
 DB TCATATTTCTACATGGTCTCTGGCTGGGAGGTTATGTTTCATCATCAACCTCATCCCT 888
 QY 275 LeuHisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyr 294
 DB CTCCATGTGTGTGTGTCTCTGATGAGAGGTACAGCAAGAGAGTCTACATAGCATAT 948
 QY 295 SerThrPhe-TyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheG 314
 DB 949 AGCAC-TTGTACATTTGGGTTTATATATATCCATGCAGATACCTTTTGGGATTCA 1007
 QY 314 nProIleArgThrSerGluHisMetAlaAlaIleValPheAlaLeuLeuGlnAlaTyr 334
 DB 1008 GCCATCAGAAACAGCAGACATGGCAGCTGCAGGTGTCTTTGCGCTGTCGCAAGCTTA 1067
 QY 334 tAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnPheGlnThrLeuPhePh 354
 DB 1068 CGCTTTTTCAGTATCTGAGACCGGTTGACAAACAGAGTTCCAGACCTTTTCTT 1127
 QY 354 eLeuGlyValSerLeuAlaIleAlaValPheLeuSerValIleTyrIleThrTyrTh 374
 DB 1128 TTTGGGTCTCTACATAGCTGAGCGCTGTGTTCTTCTAGTCTCATCTATCTGACATAC 1187
 QY 374 tGlyTyrIleAlaProTtpSerGlyArgPheTyrSerLeuTyrAspThrGlyTyrAla 394
 DB 1188 AGGTATATTGACCATGGAGTGCAGGTTTATCTCATATGAGTACTGGGTATGCAAA 1247
 QY 394 sIleHisIleProIleIleAlaSerValSerGluHisGlnProThrTyrTrpValSerPh 414
 DB 1248 AATACACATTCATATTTGATCATGAGTGTCTGAAACATCAGCCATCAGCATGGGTGTCTTT 1307
 QY 414 ePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTyrPheCysIle 434
 DB 1308 CTTCTTTGATCTACATATCTTGTATGTAACCTTCCAGAGGCTATGTTCTGCATCAA 1367
 QY 434 sAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAl 454
 DB 1368 AAATATCAACGATGAAAGAGTATTTGCTGCTGTATGCGATCAGTGTGTGTACTTTGC 1427
 QY 454 aGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAla 474
 DB 1428 CGGAGTGTATGTGGGTGATGCTGATCTGACCCCGTCTGCTGATGCTGTGCGCAT 1487
 QY 474 eAlaPheSerAsnValPheGluHisTyrLeuGlyAspMetLysArgGluAsnProPr 494
 DB 1488 CGCCTTCTCCATGTTTGTGAGCACTATTTGGGGATGACATGAAAGGAGAAACCCACC 1547
 QY 494 cValGluAspSerSerAspGluAspLysArgAsnGln-GlyAsnLeuTyrAspLysA 514
 DB 1548 TGTGGAGACAGCAGTGTATGAGGATGACAAAGAAACCC-AGGAACTTGTATGACAGG 1606
 QY 514 laGlyLysValArgLysHisAla-ThrGluGlnGluLysThr-GluGluGlyLeuGlyPr 533
 DB 1607 CAGGTAAAGTACAGGAAGCATGT-GACACAGCAAGAGAA-ACCTGAAAGAGGCTTGGGCC 1664
 QY 533 oAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHi 553
 DB 1665 CAACATCAAAAGCATTTGACCATGCTGATGCTCATGCTCTGATGATGTTTGGGTGCA 1724
 QY 553 sCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAs 573
 DB 1725 CTGACGTGGGTCAACAGCAACGCTACTCCAGTCCAAAGTGTGGTCTTGCCTCTCTACAA 1784
 QY 573 nHisAspGlyThrArgAsnIleLeuAspPheArgGluAlaTyrPheTrpLeuArgG 593
 DB 1785 TCATCATGTACACAGGATATATTAGATGATTTTGNAGAGCGTACTTTTGGCTGAGACA 1844
 QY 593 nAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaG 613
 DB 1845 AAACACGATGACACCGCCGGTCTGTGTGTGGGACTACGGGTATCAGATTTGCTGG 1904
 QY 613 yMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLe 633

DB 1905 CATGGCCAAACAGGACCACCTCTGCTGGATTAACAACACCTTGGAAACAACAGCCATCGCACT 1964
 QY 633 uValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr-LeuA 653
 DB 1965 GGTCGGAAGAGCTATCTCTTCAATGAAACGCCGCCCTATAAAATCATGAGTCC-CCTTG 2023
 QY 653 spValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleA 673
 DB 2024 ATGTCGATATGTTGGTTTATTTTCGGAGGAGTGTATGCTTATTCGGGGACGATATCA 2083
 QY 673 snLysPheLeuThrMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluS 693
 DB 2084 ACAAGTTCTCTCGATGGTTCAGATAGTGAAGGGGAGCATCCCAAGACATCCGGGAAG 2143
 QY 693 er-AspTyrPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThrLe 712
 DB 2144 G-TGACTATTTCAC-CCAGCAGGAGAGTTCCGAGTAGACAAAGCTGGTCTCTACTCT 2201
 QY 712 uLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPh 732
 DB 2202 GTTAAACTGCTTATGTATAAAATGTCTACTACAGATTTGGAGAAATGCAGCTAGATT 2261
 QY 732 eA-GThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLy 752
 DB 2262 TCGACTCCCCCAGGCTTTGACCGAACAGTAACTCTGAGATTGGAATTAAGACATTAA 2321
 QY 752 sPheLysHisLeuGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysVa 772
 DB 2322 ATTCAAGCATTTGGAGAAAGCTTTTACATCAGACAGCTGGCTTGTCCAGATATATAAGT 2381
 QY 772 lLysAlaProAspAsnArgGluThrLeu 781
 DB 2382 GAAAGACCTGACACAGGAGACACTA 2409
 RESULT 2
 CR610915 2212 bp mRNA linear HTC 21-JUL-2004
 LOCUS full-length cDNA clone CSODK010Y001 of HeLa cells Cot 25-normalized
 DEFINITION of Homo sapiens (human).
 ACCESSION CR610915
 VERSION CR610915.1 GI:50491722
 KEYWORDS HTC; CDSIT_cDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 2212)
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue
 REFERENCE 2 (bases 1 to 2212)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
 BP 131 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT 1st strand cDNA was primed with a NotI-oligo (dr) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 FEATURES
 source 1. 2212
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSODK010Y001"
 /tissue type="HeLa cells Cot 25-normalized"
 /plasmid="pCMVSPORT_6"

Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 MEDLINE
 PUBMED
 11076861
 4

REFERENCES

The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

REFERENCES

The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 5

REFERENCES

Nature 420, 563-573 (2002)
 6 (bases 1 to 2709)
 Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
 Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
 Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
 Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
 Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
 Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
 Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasabe,Y.,
 Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,D.,
 Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
 Tejima,Y., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
 Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 Direct Submission

TITLE

Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.jp,
 URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216)

COMMENT

On Dec 10, 2002 this sequence version replaced gi:12858635.
 Please visit our web site (<http://genome.gsc.riken.jp/>) for further
 details.

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5', GAGAGAGAGCGCGCACTCGAGTTTCTTTTCTTTTCTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. Second
 strand cDNA was prepared with the primer adapter of sequence [5',
 GAGAGAGAGAGGATCCAGAGCTCAATTAATTAATTAACCCGCCCC 3']. cDNA was
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
 XhoI. Host: SOLR.

FEATURES

source

1..2709
 /location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:1300006C19"
 /clone="1300006C19"
 /sex="male"
 /tissue_type="liver"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 76..2547
 /note="unnamed protein product: hypothetical
 Oligosaccharyl transferase (OTase) STT3 subunit containing
 protein (InterProIPR003674, evidence: InterPro)
 putative"
 /codon_start=1

CDS

1..2709
 /location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:1300006C19"
 /clone="1300006C19"
 /sex="male"
 /tissue_type="liver"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 76..2547
 /note="unnamed protein product: hypothetical
 Oligosaccharyl transferase (OTase) STT3 subunit containing
 protein (InterProIPR003674, evidence: InterPro)
 putative"
 /codon_start=1

/protein_id="BAB31390.1"
 /db_xref="GI:12858635"
 /translation="MSPSPAPSKSSSLNPPSWSCGLMALNSRHGHGHGPGTQASARA
 AAPPGPAGLSGSLQSDAGWQSLLSTFLFLAWLAGFSRLFAVINFESIIEFDPW
 FNYRSTHLLASHGFEFLNFWPDRAWYPLGRIVGGTVYGLMITAGLIHMLTNLIT
 VHIRDVCVFLAPTFESGLTSISTFLTRLNQAGLLAACFIAIVPGYIISRVAGSFD
 NEGIAIFALQFTYLVKTVKTSVFTWCCCLSYFYVMWAGVYFIIINLIPLHVFV
 LLLMQRYSKRVIIAYSTLYIVGLISLMQIPFVGPOPIRTSEHMAAGVAFLLQAVFL
 LLDRDLTKQEFQTLFLGLSAAAGAVFLSVIYLYTYGYIAPKSGRVSFLWDGYAKI
 HPLIASVSHQPTTWSPFDDHLVCTPAGLWFCIKMINDERVVALIYASVYF
 AGVMRLMLTLTPVCMLSAIAFNSNVEHVLGDDMKKNPVEDSDEDDKRNPNGLY
 DKAKYRKHVEQEPGEGPGLNKSIVTMLMLLMWDFYQIAGMANRTLVNDNTVN
 ASYNDGTRNLLDDFREAYFLWQNTDEHARVMSWDYGYIAGMANRTLVNDNTVN
 NSHITALYGVKMSNETAAYIKMSLDYVLYVGVGVGGDDINKFLMWVRIAGE
 HKDITREGDYFTQGGEPFVDKAGSPNTLNCMLMYKMSYRFGEGOLDPRTTGGDRTN
 AEIGNKDIKFKHLEAEFTSEHLVRIYKVKAPNDNRETGLHKPRVTNIVPKQYLSKKT
 TKRKGVVKNKLVFKGKTKSKTV"
 2693..2698
 /note="putative"
 2709
 /note="putative"
 polyA_signal
 polyA_site
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 2709
 Score: 365.00 Matches: 720
 Percent Similarity: 98.09% Conservative: 0
 Best Local Similarity: 98.09% Mismatches: 7
 Query Match: 44.19% Indels: 14
 DB: 3 Gaps: 0

US-10-028-384-2 (1-826) x AK018758 (1-2709)

QY 55 ProkaryoticLeuSerGlyLeuSerGlnProkaryoticLeuSerLeuSerPhe 74
 Db 229 CCCCAGGAGGCTGTCGCGGGGCTGTCGAGCGCGCGGAGTGGTCTCTCTTC 288
 QY 75 ThrileLeuPheLeuAlaTTPLeuAlaGlyPheSerSerArgLeuPheAlaValleArg 94
 Db 289 ACCATCTCTCTCTGCGCTGGCGCGCTTACGTCGCGCTCTCTGCGCGCTCT 348
 QY 95 PheGluSerlelleHisGluPheAspProTTPPheAsnTyrArgSerThrHisLeu 114
 Db 349 TTCGAGAGCATCATCCACGAGTTCGACCCGCTGTTAACTATAGATCAACACATCATCTT 408
 QY 115 AlaSerHisGlyPheTyrGluPheLeuAsnTTPPheAspGluArgAlaTTPyrProLeu 134
 Db 409 GCATCTCATGATCTCTATGAGTTCTTAAATTTGTTTATGAAAGAGAGATGGTACCCACTG 468
 QY 135 GlyArgileValGlyGlyThrValTyrProGlyLeuMetileThrAlaGlyLeuileHis 154
 Db 469 GGAAGAATAGTGGGTGGCCCGCTTACCCAGGCTTGTATGATACAGCTGGCTTATCAT 528
 QY 155 TTPileLeuAsnThrLeuAsnilleThrValHisileArgaspValCysValPheLeuAla 174
 Db 529 TGGATTTTAAATACATTGAACATACAGATTACATAGAGATGTGTGTATTCTTCTTGA 588
 QY 175 ProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTTPAsn 194
 Db 589 CCAACTTTTAGCGGCTTATCATCATCATCTACGCTTCTGCTACTAGAGAGACTGTGGAAC 648
 QY 195 GlnGlyAlaGlyLeuLeuAlaAlaCysPheilleAlaileValProGlyTyrileSerArg 214
 Db 649 CAAGGAGCAGAGACTTCTAGCTGCTTGTCTTCAATGCTATGCTATGCTATGCTATGCT 708
 QY 215 SexValAlaGlySerPheAspAsnGluGlyIleAlailePheAlaLeuGlnPheThrTyr 234
 Db 709 TCAGTGGCGGAGTCCCTTGTATATGAGGATTCGCTATGCTATGCTATGCTATGCTATGCT 768
 QY 235 TyrileTTPVallySerVallySerVallySerValPheThrMetCysCysLeu 254
 Db 769 TACTTATGGTAAAGTCTGTGAAGACCGGCTCTGTCTCTGAGACATGTGCTGCTGCTGCT 828

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:58 ; Search time 6394.05 Seconds
(without alignments)
4707.376 Million cell updates/sec

Title: US-10-028-384-2
Perfect score: 826
Sequence: 1 MAEPAPSSKHKSLNSPPW.....GYIKNLVFKKGKIKKTV 826

Scoring table:

OLIGO	Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 32822875 seqs, 18219865908 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004.131646_6807/app_query.fasta_1.3740
-DB=EST -QFMT=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.edi -LIST=45
-DOCLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US10028384 @CNG 1.13960 @runat_14122004.131646_6807 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_hc1:*
- 4: gb_est3:*
- 5: gb_est4:*
- 6: gb_est5:*
- 7: gb_est6:*
- 8: gb_gss1:*
- 9: gb_gss2:*

Result No.	Score	Query Match	Length	ID	Description
1	365	44.2	2709	AK018758	AK018758 Mus muscu
2	282	34.1	2212	3 CR610915	CR610915 full-len
3	243	29.4	2669	3 AK012153	AK012153 Mus muscu
4	240	29.1	852	7 CN458732	CN458732 UI-M-HBO
5	235	28.5	747	2 BE952197	BE952197 UI-M-CDO
6	232	28.1	775	7 CK032376	CK032376 AGENCOURT
7	230	27.8	693	7 CN461211	CN461211 UI-M-HNO
8	227	27.5	950	5 BU515954	BU515954 AGENCOURT
9	219	26.5	658	6 CA511417	CA511417 UI-R-FJO

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc1:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ALIGNMENTS

2709 bp mRNA linear HTC 03-APR-2004

AK018758 Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300006C19 product:hypothetical Oligosaccharyl transferase (OTase) STT3 subunit containing protein, full insert sequence:

AK018758 GI:26384577

AK018758.2 HTCC, CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1. Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

2. Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

3. Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,


```

Db      62 ATAGCTGGATGGCTAATAGAACTACCTGGTGGATAAACAACCTGGGATAACAGCCAC 121
Qy      631 IleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArg 650
Db      122 ATAGCACTGGTGGGAAAGATATGCTCTCTTAATGAACAGCAGCCCTATAAAATCATGAGG 181
Qy      651 ThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAsp 670
Db      182 ACTCTAGATGATAGATATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGGTAT 241
Qy      671 AspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIle 690
Db      242 GATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAGGAGAGAACATCCCAAGACATT 301
Qy      691 ArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerPro 710
Db      302 CGGGAAGTGACTATTTTACCCACAGGAGGAATTCGGTGTAGACAAAGCAGGATCCCT 361
Qy      711 ThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeu 730
Db      362 ACTTTGTGAATGCTTATGTATAAATGTCATACAGATTTGGAGAAATGCAGCTG 421
Qy      731 AspPheArg 733
Db      422 GATTTTCGT 430

```

RESULT 15

```

US-10-046-935-717
; Sequence 717, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yudiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-046-935-717

```

```

Alignment Scores:
Pred. No.:      3,17e-141      Length:      430
Score:          143.00      Matches:      143
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      17.31%      Indels:      0
DB:              13      Gaps:      0

```

US-10-028-384-2 (1-826) x US-10-046-935-717 (1-430)

```

Qy      591 LeuArgGlnAsnThrAspGluHisAlaArgValMetSerTyrTyrAspTyrGlyTyrGln 610
Db      2 CTAAGGCAAAATACAGATGAACATGCACGAGTAATGCTCTGGTGGGATTTATGGCTATCAG 61
Qy      611 IleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnThrTyrAsnAsnSerHis 630
Db      62 ATAGCTGGATGGCTATAGACTAGCTGTTGGTGGATAAACAACCTGGGATAACAGCCAC 121
Qy      631 IleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArg 650
Db      122 ATAGCACTGGTGGGAAAGATATGCTCTCTTAATGAACAGCAGCCCTATAAAATCATGAGG 181
Qy      651 ThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAsp 670

```

```

Db      182 ACTCTAGATGATAGATATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGGTAT 241
Qy      671 AspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIle 690
Db      242 GATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAGGAGAGAACATCCCAAGACATT 301
Qy      691 ArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerPro 710
Db      302 CGGGAAGTGACTATTTTACCCACAGGAGGAATTCGGTGTAGACAAAGCAGGATCCCT 361
Qy      711 ThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeu 730
Db      362 ACTTTGTGAATGCTTATGTATAAATGTCATACAGATTTGGAGAAATGCAGCTG 421
Qy      731 AspPheArg 733
Db      422 GATTTTCGT 430

```

Search completed: December 15, 2004, 15:26:48
Job time : 1072.95 secs

Qy 630 HisleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMet 649
Db 423 CACATAGCACTGGTGGAAAGCTATGCTCTTAATGAACACAGCAGCCTATAAAATCATG 482
Qy 650 ArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 669
Db 483 AGNACTCTAGATGAGTAAATGTTTGGTATTTTGGAGGGTATTTGGCTATTTCTGGT 542
Qy 670 AspAspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAsp 689
Db 543 GATGATATCAACAAATTTCTCTCGATGGTTAGGATAGCTGAAGGAGAACATCCCAAGAC 602
Qy 690 IleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySer 709
Db 603 ATTCCGGAAGTAGTATTTTACCCACAGGGGAGATTTCCGTGTAGACAAAGCAGGATCC 662
Qy 710 ProThrLeuLeu-AsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetG1 729
Db 663 CCTACTTTGKT-GAATTCCTTATGATTAATAATGTCATACTACAGATTTGGAGAAATGCA 721
Qy 729 nLeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLys 749
Db 722 GCTGGATTTCTACACCCCGAGTTTGACCGAACACGTAATGCTGAGATTTGGAAATAA 781
Qy 749 sAspIleGlyPheLysHisLeuGluGluAlaPheThrSerGluHisTyrLeuValArgI1 769
Db 782 GGACATTAATTAACACATTTGAAGAAGCCCTTTACATCAGAACACACTGGCTTTGTAGGAT 841
Qy 769 eTyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAs 789
Db 842 ATATAAGTAAGACACCTGATTAACAGGGAGACATTAGATCACAACCTCGAGTCACCAA 901
Qy 789 nilePheProGlnLysTyrLeuSerLysLysThrLysArgLysArgGlyTyrI1 809
Db 902 CATTTTCCCAACAGAGTATTTGTCAAGAAGACTACCAAAAGAAAGCGGTGGCTACAT 961
Qy 809 eLysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLysThrVal 826
Db 962 TAAAAATAGCTGGTTTAAAGAAAGCAGAAATATCTAAGAGACACTGTT 1013

RESULT 13

US-09-998-598-1643
; Sequence 1643, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1643
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1643

Alignment Scores:
Pred. No.: 1,638-165 Length: 500
Score: 166.00 Matches: 166
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.10% Indels: 0
DB: 9 Gaps: 0

US-10-028-384-2 (1-826) x US-09-998-598-1643 (1-500)

Qy 158 AsnThrLeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPhe 177
Db 2 AATACATTGAACATAAATGTTTACATAAGAGACGCTATGTTGTTCTTCCCTTGACCAACTTTT 61
Qy 178 SerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTyrAsnGlnGlyAla 197
Db 62 AGGGGCTTTACATCTATATCTACTTTCTCTTACAGAGAACTTTGGAACCAAGGAGCA 121
Qy 198 GlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAla 217
Db 122 GGAATTTTACGCTGCTTTTATTTATGCTATTGTACAGGCTACATCTCGTTCAGTAGCT 181
Qy 218 GlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyr 237
Db 182 GGATCTCTTGAATAAGAGGCAATTCATTTTGGCACTTCAGTTCCACATCTATTTATGG 241
Qy 238 ValLysSerValLysThrGlySerValPheThrMetCysCysLeuSerTyrPhe 257
Db 242 GTAAATCTGTAAAACTGGGTCAGTTTTTTGGACAATGTCTGTCTGCTTATCTCTATTTTC 301
Qy 258 TyrMetValSerAlaTyrGlyTyrValPheIleIleAsnLeuIleProLeuHisVal 277
Db 302 TATATGCTCTGCTTGGGGTGGTTATGTAATTTATCATCAATCTTATTCACATCGATGTA 361
Qy 278 PheValLeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPhe 297
Db 362 TTGTGTTTGTACTGATGAGAGATACAGCAAGAGCTACATAGCATATAGCACTTTC 421
Qy 298 TyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPheGlnProIleArg 317
Db 422 TACATTTGCTGGTTTAAATTAATTAATCAATGCAGATACCTTTTGGGATTCACGCCAATCAGA 481
Qy 318 ThrSerGluHisMetAla 323
Db 482 ACAAGTGAACACATCGCA 499

RESULT 14

US-09-878-178-717
; Sequence 717, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuguu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-878-178-717

Alignment Scores:
Pred. No.: 3,178-141 Length: 430
Score: 143.00 Matches: 143
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.31% Indels: 0
DB: 9 Gaps: 0

US-10-028-384-2 (1-826) x US-09-878-178-717 (1-430)

Qy 591 LeuArgGlnAsnThrAspGluHisAlaArgValMetSerTyrTyrAspTyrGlyTyrGln 610
Db 2 CTAAGGCAAAATACAGATGAACATGCAGGAGTAATGTCTTGTGGGATTTATGGCTATCAG 61
Qy 611 IleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnThrTyrPheAsnSerHis 630

```

RESULT 11
US-10-296-115-629
; Sequence 629, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: NO. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 629
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-629

Alignment Scores:
Pred. No.: 3 51e-219 Length: 1114
Score: 217.00 Matches: 274
Percent Similarity: 98.21% Conservative: 0
Best Local Similarity: 98.21% Mismatches: 2
Query Match: 26.27% Indels: 5
DB: 16 Gaps: 0

US-10-028-384-2 (1-826) x US-10-296-115-629 (1-1114)
QY 1 MetAlaGluProSerAlaProGluSerHisHisSerSerLeuAsnSerSerProTrrp 20
Db 211 ATGGCGAGCCCTCGGCCCGGAGAGCAAGCAAGTCGTCCTCAATCTGCCCGGG 270
QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
Db 271 AGTGCCCTCATGGCCCTGGAAACAGCGCGCCGCCACCGCGCGCGCGCGCGCG 330
QY 41 AlaHisLysAlaAlaGlyAlaAlaProProLysProAlaPro---AlaGlyLeuSer 59
Db 331 GCGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTG-CTCAC- 388
QY 60 GlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeu 79
Db 389 GGGGGGCTGTGCGAGCGCGCTGGTGGCAGTCGCTTCCTTCCTTCATCTCTCTCTG 448
QY 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIlele 99
Db 449 GCCTGGCTTGGCGGCTTCAGCTCGCGCTTCCTTCGCGCTTCATCGCTTCGAAAGCATATC 508
QY 100 HisGluPheAspProTrpPheSerArgSerArgSerArgSerArgSerArgSerArgSer 119
Db 509 CACGAGTTCACCGCGGTGTTAACTATAGATCAACATATCTTCCTTCATCTCATCGGTC 568
QY 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGly 139
Db 569 TATGAATTTTAAATTTTGTATGAAGAGCATGATATCCACTAGGAAGATAGTAGT 628
QY 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
Db 629 GGTACTGTTTACCAGCGGTGATGATAACCGCTGGCTTATTCATTGGATTTTAAATACA 688
QY 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
Db 689 TTGACATACTGTTACATACAGACGATATGCTGCTTCCTTCACCACTTTTACGCGC 748
QY 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199
Db 749 CTTACATCTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 808
QY 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219

```

```

Db 809 TTAGCTGCTGTTTATTGCTATTGTACACGCTACATATCTCGTCACTAGTGGATCC 868
QY 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrLeuTrpVallys 239
Db 869 TTTGATAATGAAGGCATTCCTATTGTCACCTTCAGTTTCACATCTATTTATGGGTAATA 928
QY 240 SerValIleThrGlySerValPheTrpThrMetCysCysCysLeuSerTyrPheTyrMet 259
Db 929 TCTGTAAAAACCTGGGTCACTTTTGGCAATGCTGCTGCTTATCTATTTCTATATG 988
QY 260 ValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHis 276
Db 989 GTCTCTGCTTGGGGTGGTATGATTTATCATCAATCTTATTCACCTGCAT 1039

RESULT 12
US-10-264-049-630
; Sequence 630, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Barse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 630
; LENGTH: 2660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2634)..(2634)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2637)..(2637)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-630

Alignment Scores:
Pred. No.: 1 37e-196 Length: 2660
Score: 196.00 Matches: 296
Percent Similarity: 99.33% Conservative: 0
Best Local Similarity: 99.33% Mismatches: 1
Query Match: 23.73% Indels: 2
DB: 16 Gaps: 0

US-10-028-384-2 (1-826) x US-10-264-049-630 (1-2660)
QY 530 GlyLeuGlyProAsnIleIleValThrMetLeuMetLeuMetMetMet 549
Db 123 GGGTAGCCCTAATATAAAGCATTGTCACCATGTTGATGCTGATGATGATG 182
QY 550 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeu 569
Db 183 TTTGCTGTCCACTGTACTGGGTCAAGCAATGCTTCTAGTCCAAAGTAGTCCCTG 242
QY 570 AlaserTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPhe 589
Db 243 GCCTCAATCAATCATGATGACCCAGGATATCTTATGATGATTTTAGAGAGCTTACTTT 302
QY 590 TrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyr 609
Db 303 TGGCTAAGGCAATAACAGATGAACATGACAGCATGATGTTGGTGGGATTATGGCTAT 362
QY 610 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnSer 629
Db 363 CAGATAGTGAATGGTGAATAGAACTACGTTGGTGGTGAATATACACCTGGAAATACAGC 422

```


Db 242 TGTGTGTTCTTCTGACCAACTTTTAGGGCTTTACATCTATATCTACTTTCTGCTTACA 301
 Qy 190 ArgGluLeuThrPheAsnGlnGlyAlaGlyLeuLeuAlaCysPheLeuAlaValPro 209
 Db 302 AGAAGAACTTTGGAACCAAGGACGAGGACTTTTAGCTGCTGTGTTTATTCTATTGTACA 361
 Qy 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 Db 362 GGCTACATATCTCGTCTAGTAGTGGATCTTTGATATAGAGGCATCTCTATTGTGA 421
 Qy 230 LeuGlnPheThrTyrTyrLeuThrValIleValSerValIleThrGlySerValPheThr 249
 Db 422 CTTGAGTTCATATCTATTATGGTAAATCTGTAAACCTGGGTCAGTTTGTGGACA 481
 Qy 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaThrGlyGlyTyrValPheIle 269
 Db 482 ATGTGCTGCTCTTATCTCTATCTATGCTCTCTGCTGGGTGTTATGATTATC 541
 Qy 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnThrGlySerIleArg 289
 Db 542 ATCAATCTTATCTCACTGATGATTTGTGTTGTTTGTACTGATGACAGATACAGAAAAGA 601
 Qy 290 ValTyrIleAlaTyrIleSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
 Db 602 GTCTACATAGCATATAGCACTTTCTACATTTGTGGTGTAAATATTATCAATGCAGATACCT 661
 Qy 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
 Db 662 TTGTGGGATTCACGCAACATCAGAACAGTGAACATGCGAGCTGCGAGTGTCTTTGCA 721
 Qy 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrIleGlnPhe 349
 Db 722 TTGTGCAAGCTTATGCTTTCTTCTGAGTATCTGAGAGCCGATTAACAAAACAAGAGTTC 781
 Qy 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
 Db 782 CAGACCCCTTTCTTTTGGGTGATCTACTAGCTGCGAGGTGCTGTGTTCTTATGTGATC 841
 Qy 370 TyrLeuThrTyrThrGlyTyrIleAlaProTyrPheTyrGlyArgPheTyrSerLeuThrPasp 389
 Db 842 TATTGTACTTATACAGCTTACATTCGACCATGCGATGCGAGTGGCAGGTTTATTCTATTGGGAT 901
 Qy 390 ThrGlyTyrAlaIleIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
 Db 902 ACTGGGTATGCAAAATACATCTCCAAATATTGCAATCATGCTGTCTGAGCATCAACCTACG 961
 Qy 410 ThrTyrValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
 Db 962 ACTTGGGTGCTTTCTTCTTCTTGTATCTATATCTTGTATGATCTTCCAGCAGGCTT 1021
 Qy 430 TrpPheCysIleLeuAsnIleAsnAspGluArg 440
 Db 1022 TGGTTCTGCAATCAAAATATCAACGATGAAGA 1054

RESULT 9

US-09-818-683-133
 ; Sequence 133, Application US/09818683
 ; Publication No. US20040185440A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Feng et al.
 ; TITLE OF INVENTION: 125 Human Secreted Proteins
 ; FILE REFERENCE: P2020P1
 ; CURRENT APPLICATION NUMBER: US/09/818,683
 ; CURRENT FILING DATE: 2001-03-28
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 612
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 133
 ; LENGTH: 1543
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE

; LOCATION: (1055)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (1143)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; US-09-818-683-133
 Alignment Scores:
 Pred. No.: 0 Length: 1543
 Score: 351.00 Matches: 351
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 42.49% Indels: 0
 DB: 11 Gaps: 0

US-10-028-384-2 (1-826) x US-09-818-683-133 (1-1543)

Qy 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArg 109
 Db 2 TTGCGCGTTCATCCGCTTCGAAGCATCATCCACGAGTTCGACCCGTGGTTAACTATAGA 61
 Qy 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTyrPheAspGluArg 129
 Db 62 TCAACACATCATCTTGCACTTCATGGTCTCTATGAATTTTAAATTTGGTTGATGAAGA 121
 Qy 130 AlaTyrTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
 Db 122 GCATGTATTCCTACTAGGAAGATAGTAGTGGTACTGTTTACCAGGCTTGATGATAACC 181
 Qy 150 AlaGlyLeuIleHisTyrPheLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
 Db 182 GCTGGCTTATTTCATTGGATTTTAAATACATTTGAACATACTGTTACATAAGAGACGTA 241
 Qy 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 Db 242 TGTGTGTTCTTCTGCAACACTTTTAGCGGCTTACATCTATATCTATCTTTCTGCTTACA 301
 Qy 190 ArgGluLeuThrPheAsnGlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValPro 209
 Db 302 AGAAGAACTTTGGAACCAAGGACGAGGACTTTTAGCTGCTGTTTATTGCTATTGTACCA 361
 Qy 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 Db 362 GGCTACATATCTCGTCTAGTAGTGGATCTTTGATATAGAGGCATCTGCTATTTTTGTGA 421
 Qy 230 LeuGlnPheThrTyrTyrLeuThrValIleValSerValIleThrGlySerValPheThr 249
 Db 422 CTTGAGTTCATATCTATTATGGTAAATCTGTAAACCTGGGTGCTGCTTTTGTGACA 481
 Qy 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaThrGlyGlyTyrValPheIle 269
 Db 482 ATGTGCTGCTCTTATCTCTATGATGCTCTGCTGGGTGTTATGATTATTC 541
 Qy 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerIleArg 289
 Db 542 ATCAATCTTATCTCACTGATGATTTGTGTTGTTTGTACTGATGACAGATACAGAAAAGA 601
 Qy 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
 Db 602 GTCTACATAGCATATAGCACTTTCTACATTTGTGGTGTAAATATTATCAATGCAGATACCT 661
 Qy 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
 Db 662 TTGTGGGATTCACGCAACATCAGAACAGTGAACACATGCGAGCTGCGAGTGTCTTTTGA 721
 Qy 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrIleGlnPhe 349
 Db 722 TTGTGCAAGCTTATGCTTTCTTCTGAGTATCTGAGAGCCGATTAACAAAACAAGAGTTC 781
 Qy 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
 Db 782 CAGACCCCTTTCTTTTGGGTGATCTACTAGCTGCGAGGTGCTGTGTTCTTATGTGATC 841

```

; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-305-736-133

Alignment Scores:
Pred. No.: 0 Length: 1543
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x US-09-305-736-133 (1-1543)

QY 90 PheAlaValIleArgPheGluSerIleHisGluPheAspProTrpPheAsnTyrArg 109
DB 2 TTGCGCGTCATCCGCTTCGAAGCATCATCCAGAGTTCGACCGCGTGGTTAACTATAGA 61
QY 110 SerThrHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
DB 62 TCAACACATCATCTTGCACTCATCGGTTCTATGAATTTTAAATGGTTGATGAAGA 121
QY 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
DB 122 GCATGGTATCCACTAGGAAGAATAGTAGGTGCTCTTTACCCAGCGGTTGATGAACC 181
QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
DB 182 GCTGCGCTTATTCATCGATTTTAAATACATTCGACATACTGTTACATGAGACGTA 241
QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuThr 189
DB 242 TGTGTGTTCTTGCACCAACTTTTAGCGGCTTACATCTATATCTACTTTCTCTGCTTAC 301
QY 190 ArgGluLeuTrpAsnGlnGlyValGlyLeuAlaAlaCysPheIleAlaIleValPro 209
DB 302 AGAGACCTTTGGACCAAGAGAGAGAGACTTTAGCTGCTGTTTATTGCTATTGATGACA 361
QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
DB 362 GGCTACATATCTCGCTCAGTAGCTGGATCCTTTGATAATGAAGGCATTCGTATTTGCA 421
QY 230 LeuGlnPheThrTyrTyrLeuTrpValIleSerValIleThrGlySerValPheTrpThr 249
DB 422 CTTCACTTACATACATATTATGGGTAAATCTGTAAATACTGGGTCAGTTTTCGACA 481
QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
DB 482 ATGTGCTGCTGCTTATCTCTATTTCTATATGCTCTCTGCTTGGGTTGTTATGTTATTC 541
QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArg 289
DB 542 ATCAATCTTATTCACATGCATGATTTTGTGTTACTGATGACAGATACACGACAAAGA 601
QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
DB 602 GTCTACATAGCATATAGCACTTTCTACATTTGTGGGTTTAAATATTATTCATGACAGATCCT 661
QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
DB 662 TTTGTGGGATTCAGCCCAATCAGAACAGTGAACACATGCGACCTGCAGGTGCTTTGCA 721
QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgIleThrLysGlnGluPhe 349
DB 722 TTGCTGCAAGCTTATGCTTTCTTCAGTATCTGAGACCGGATTAACAAACAGAGTTC 781
QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369

```

```

DB 782 CAGACCCCTTTCTTTTGGGTGTATCAGTAGCTGAGGCTGTGTCTTCTTAGTGTATC 841
QY 370 TyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAsp 389
DB 842 TATTGTGACTTATACAGTTACATTCACCATGAGTGGCAGGTTTTTATTCAITGTGGAT 901
QY 390 ThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
DB 902 ACTGGGTATGCAAAATAACACATTCCTCAATTTATGATCAGTCTCTGAGCATCAACTACG 961
QY 410 ThrTrpValSerPhePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
DB 962 ACTGGGTGCTCTTCTCTTTGATACATATTTCTGTATGTACCTTCCACAGCGCCTT 1021
QY 430 TrpPheCysIleLysAsnIleAsnAspGluArg 440
DB 1022 TGGTTCCTGCATCAAAAATATCAACGATGAAGA 1054

RESULT 8
US-09-818-683-133
; Sequence 133, Application US/09818683
; Publication No. US20030211472A1
; GENERAL INFORMATION:
; APPLICANT: Peng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/818,683
; CURRENT FILING DATE: 2001-03-28
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-818-683-133

Alignment Scores:
Pred. No.: 0 Length: 1543
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x US-09-818-683-133 (1-1543)

QY 90 PheAlaValIleArgPheGluSerIleHisGluPheAspProTrpPheAsnTyrArg 109
DB 2 TTGCGCGTCATCCGCTTCGAAGCATCATCCAGAGTTCGACCGCGTGGTTAACTATAGA 61
QY 110 SerThrHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
DB 62 TCAACACATCATCTTGCACTCATCGGTTCTATGAATTTTAAATGGTTGATGAAGA 121
QY 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
DB 122 GCATGGTATCCACTAGGAAGAATAGTAGGTGCTCTTTACCCAGCGGTTGATGAACC 181
QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
DB 182 GCTGCGCTTATTCATCGATTTTAAATACATTCGACATACTGTTACATGAGACGTA 241
QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuThr 189
DB 242 TGTGTGTTCTTGCACCAACTTTTAGCGGCTTACATCTATATCTACTTTCTCTGCTTAC 301
QY 190 ArgGluLeuTrpAsnGlnGlyValGlyLeuAlaAlaCysPheIleAlaIleValPro 209
DB 302 AGAGACCTTTGGACCAAGAGAGAGAGACTTTAGCTGCTGTTTATTGCTATTGATGACA 361
QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
DB 362 GGCTACATATCTCGCTCAGTAGCTGGATCCTTTGATAATGAAGGCATTCGTATTTGCA 421
QY 230 LeuGlnPheThrTyrTyrLeuTrpValIleSerValIleThrGlySerValPheTrpThr 249
DB 422 CTTCACTTACATACATATTATGGGTAAATCTGTAAATACTGGGTCAGTTTTCGACA 481
QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
DB 482 ATGTGCTGCTGCTTATCTCTATTTCTATATGCTCTCTGCTTGGGTTGTTATGTTATTC 541
QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArg 289
DB 542 ATCAATCTTATTCACATGCATGATTTTGTGTTACTGATGACAGATACACGACAAAGA 601
QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
DB 602 GTCTACATAGCATATAGCACTTTCTACATTTGTGGGTTTAAATATTATTCATGACAGATCCT 661
QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
DB 662 TTTGTGGGATTCAGCCCAATCAGAACAGTGAACACATGCGACCTGCAGGTGCTTTGCA 721
QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgIleThrLysGlnGluPhe 349
DB 722 TTGCTGCAAGCTTATGCTTTCTTCAGTATCTGAGACCGGATTAACAAACAGAGTTC 781
QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369

```

```

; PRIOR APPLICATION NUMBER: US 60/066,090
; PRIOR FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a.t.g. or c
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a.t.g. or c
US-09-974-879-133

Alignment Scores:
Pred. No.: 0 Length: 1543
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x US-09-974-879-133 (1-1543)

QY 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 109
DB 2 TTGCGCGTCATCCGCTTCGAAAGCATCATCCAGAGTTCGACCCGGTGGTTAACTATAGA 61
QY 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
DB 62 TCAACACATCATCTTGCATCTCTGCTGCTCTATGCAATTTTAAATGGTTGATGAAAGA 121
QY 130 AlaTrpTyrProLeuGluValGileValGlyGlyThrValTyrProGlyLeuMetIleThr 149
DB 122 GCATGGTATCCACTAGGAGAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATACC 181
QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
DB 182 GCTGCGCTTATTCATTCGATTTTAAATACATTAACATTAACATTAACATTAACATTAAC 241
QY 170 CysValPheLeuAlaProTrpPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
DB 242 TGTGTGTCCTTGCACCACTTTTAGCGGCTTACATCATATCTACTTTCTGCTTACA 301
QY 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValPro 209
DB 302 AGAAGAACTTTGGAACCAAGGACGAGGACTTTTAGCTGCTGTTTATTCCTATTGTACCA 361
QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
DB 362 GGCTACATATCTCGTCAGTAGCTGGATCCTTTGATAATCAAGGCATTCCTATTITGCA 421
QY 230 LeuGlnPheThrTyrTyrLeuTrpValIleSerValIleThrGlySerValPheTrpThr 249
DB 422 CTTGAGTTCACATACTATTATGGGTAAATCTGTAAACCTGGGTCAGCTTTTGGACA 481
QY 250 MetCysCysLeuSerTyrPheTrpMetValSerAlaTrpGlyGlyTyrValPheIle 269
DB 482 ATGTGCTGCTCTTATCTATCTATCTATGCTCTGCTGGGTGTTATGTTATTC 541
QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerIleArg 289
DB 542 ATCAATCTTATCCACGTCATGATTTTGTGTTACTGATGACAGATACAGCAAAAGA 601
QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuLeuSerMetGlnIlePro 309
DB 602 GTCTACATAGCATATAGCACTTTCTACATGTTGGGTTTAAATTAATCAATGACAGATACCT 661
QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329

```

```

DB 662 TTTGTGGGATCCAGCCATCAAGCAAGTGAACACATGGCAGCTGCAGGTGCTTTTGA 721
QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrIleGlnGluPhe 349
DB 722 TTGCTGCAAGCTTATGCTTTTCTTGCAGTATCTCGAGAGCCGATTAACAAAAACAAGAGTTC 781
QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
DB 782 CAGACCCCTTTCTTTTGGGTGTATCACTAGCTGACGGTGTGCTGCTGCTTCTAGTGTATC 841
QY 370 TyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAsp 389
DB 842 TATTTGACITATACAGGTTACATTCGACCATCGAGTGGCAGGTTTATTTATTCATTGTTGGAT 901
QY 390 ThrGlyTyrAlaIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
DB 902 ACTGGGTATGCAAAATACATTCATATTATGATCAGTGTCTGAGCANTCAACCTACG 961
QY 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
DB 962 ACTTGGGTGCTTTCTTCTTGTATCTACATATTTCTGTATGTACCTTCCACAGAGGCTT 1021
QY 430 TrpPheCysLeuValAsnIleAsnAspGluArg 440
DB 1022 TGGTTCTGCATCAAAATATCAACGATGAAGA 1054

```

RESULT 7

```

US-09-305-736-133
; Sequence 133, Application US/09305736
; Publication No. US20030088078A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/305,736
; EARLIER FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: PCT/US98/23435
; EARLIER FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/064,911
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,912
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,983
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,900
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,988
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,987
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,908
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,984
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,985
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/066,094
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,100
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,089
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,095
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,090
; EARLIER FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA131PI

; CURRENT APPLICATION NUMBER: US/10/264,237

; CURRENT FILING DATE: 2002-10-04

; PRIOR APPLICATION NUMBER: PCT/US01/16450

; PRIOR FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: US 60/205,515

; PRIOR FILING DATE: 2000-05-19

; NUMBER OF SEQ ID NOS: 2876

; SOFTWARE: PatentIn Ver. 3.1

; SEQ ID NO 412

; LENGTH: 1209

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1203)..(1203)

; OTHER INFORMATION: n equals a,t,g, or c

US-10-264-237-412

Alignment Scores:

Pred. No.: 0 Length: 1209
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DB: 16 Gaps: 0

US-10-028-384-2 (1-826) x US-10-264-237-412 (1-1209)

QY 90 PheAlaValIleAArgPheGluSerIleHisGluPheAspProTrpPheAsnTrpArg 109
Db 2 TTCGGCGTATCCGGTTCGAAGCANTATCCACGAGTTCACCGTGGTTAACTATAGA 61
QY 110 SerThrHisHisLeuAlaSerHisGlyPheTrpGluPheLeuAsnTrpPheAspGluArg 129
Db 62 TCAACACATCATCTTGCATCTCATGGGTCTATGAATTTTAAATGGTTGATGAAGA 121
QY 130 AlaTrpTrpProLeuGlyArgIleValGlyGlyTrpValTrpProGlyLeuMetIleThr 149
Db 122 GCATGATATCCCTAGGAAGATAGTAGTGGTACTGTTACCCAGGTTGATGAAC 181
QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleAArgAspVal 169
Db 182 GCTGGCCTTATTCATGGATTTAAATACATTAACATTAACATTAACATTAACATTA 241
QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
Db 242 TGTGTGTCCTTGCACCACTTTAGCGGCTTACATCTATATCTACTTTCCCTGCTTACA 301
QY 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValPro 209
Db 302 AGAGAACTTGAACCAAGGACGAGACTTTTAGCTGCTTTTATTTGCTATTGTACCA 361
QY 210 GlyTrpIleSerArgSerValAlaGlySerPheAspGlnGlyIleAlaIlePheAla 229
Db 362 GGTACATATCTCGGTAGTACTGATGATCTTTGATTAAGGCAATGCTATTATTGCA 421
QY 230 LeuGlnPheThrTrpTrpTrpValysSerValysThrGlySerValPheTrpThr 249
Db 422 CTTCACTTACATACATTAATTTAGTAAATCTGTAATCTGATGATGATGATGATGAT 481
QY 250 MetCysCysLeuSerTrpPheTrpMetValSerAlaTrpGlyTrpValPheIle 269
Db 482 ATGTGCTGCTGCTTATCTCTATATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 541
QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTrpSerLysArg 289
Db 542 ATCAATCTTATTCCTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 601
QY 290 ValTrpIleAlaTrpSerTrpPheTrpIleValGlyLeuIleLeuSerMetGlnIlePro 309
Db 602 GTCTACATAGCATATAGCACTTCTACATTTGCTGCTGCTGCTGCTGCTGCTGCT 661

QY 310 PheValGlyPheGlnProIleAArgThrSerGluHisMetAlaAlaAlaGlyValPheAla 329
Db 662 TTTGTGGGATTCCAGGCAATCAGAACAGTGAACACATGCAGCTGCAGGTCTCTTTGCA 721
QY 330 LeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeuAspArgLeuThrLysGlnGluPhe 349
Db 722 TTGCTGCAAGCTTATGCTTTCTTCAGATATCTGAGACCGATTAAACAAAACAGATTC 781
QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
Db 782 CAGACCCCTTTCTTTTGGGTGATCACTAGCTGCAGGTGCTGTCTTCTAGTGTCTATC 841
QY 370 TyrLeuThrTrpThrGlyTyrIleAlaProTrpSerGlyArgPheTrpSerLeuTrpAsp 389
Db 842 TATTGACTTATACAGATTACATTCACCATGGAGTGGCAGGTTTATTTCATTGGGAT 901
QY 390 ThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
Db 902 ACTGGGTATGCAAAAATACACATTTCCAATTTATTCATCATCTATCTTGTATGATACCT 961
QY 410 ThrTrpValSerPhePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
Db 962 ACTTGGGTGCTTTCTTTTGTATCTACATATCTTGTATGATACCTTCCACAGCGCTT 1021
QY 430 TrpPheCysIleLysAsnIleAsnAspGluArg 440
Db 1022 TGGTTCTGCATCAAAAATATCAACGATGAAGA 1054

RESULT 6

US-09-974-879-133
; Sequence 133, Application US/09974879
; Publication No. US20030028003A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P2
; CURRENT APPLICATION NUMBER: US/09/974, 879
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/239, 893
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/818, 683
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 09/305, 736
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: PCT/US98/23435
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/064, 911
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 912
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 983
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 900
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 988
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 987
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 908
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 984
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 985
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/066, 094
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/056, 100
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066, 089
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066, 095
; PRIOR FILING DATE: 1997-11-17

670	ATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAGGAGAACATCCCAAGACATTCCG	729	Db
692	GlusSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThr	711	Qy
730	GAAAGTGACTATTTATCCCCACAGGAGAAATCCGTGTAGACAAAGCAGGATCCCCTACT	789	Db
712	LeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAsp	731	Qy
790	TTGTGTAATTCGCTTATGTATAAAATGTGCATACTACAGATTGGAGAAATGCAGCTGAT	849	Db
732	PheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIle	751	Qy
850	TTTTCGTACACCCCGAGTTTGACCCAGACACGTAAATGCTGAGATTGGAAATAAGACATT	909	Db
752	LysPheLysHisLeuGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLys	771	Qy
910	AAATTCAACATTTGGAAGAAGCGCTTTACATTCAGAACACTGGCTTGTAGGATATATAA	969	Db
772	VallysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAsnIlePhe	791	Qy
970	GTAAGAAGCCTGTATACAGGGAGACATTAGATCAAAACCTCCAGTCAACAACATTTC	1029	Db
792	ProLysGlnLysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIleLysAsn	811	Qy
1030	CCAAAAAGAGAGTATTGTTCAAAGACAGACTACAAAGGAAGCGTGGCTACATTAAAAT	1089	Db
812	LysLeuValPheLysLysGlyLysLysIleSerLysLysThrVal	826	Qy
1090	AAGCTGGTGTTTTAAGAAGGCAAGAAATATCTAAGAAGACTGTT	1134	Db

RESULT 4

```

Record 4
US-10-106-698-330
; Sequence 330, Application US/101066598
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005PL
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 330
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1203)..(1203)
; OTHER INFORMATION: n equals a.t.g. or c
US-10-106-698-330

```

Db 1369 AAATATCAACGATGAAAGAGATTTTGGCTCTGATGCGATCAGTGCTGTGACTTTGC 1428
QY 454 aGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAla1 474
Db 1429 CGAGTGTATGTTGGCTGATGCTCACTCTCACCCCGTGTCTGTGATGCTGTGCGCCAT 1488
QY 474 eAlaPheSerAsnValPheGluHisLeuGlyAspMetLeuValArgGluAsnProPr 494
Db 1489 CGCTTCTCCATGTTTGGAGCACTATTGGGGGATGACATGAAAGGGAAACCCACC 1548
QY 494 oValGluAspSerSerAspGluAspAspLysArgAsnGln-GlyAsnLeuTyrAspLysA 514
Db 1549 TGTGGAGGACACGACGTGATGAGGATGACAAAGAAACCC-AGGAAACTTGTATGACAAG 1607
QY 514 laGlyLysValArgLysHisAla-ThrGluGlnGluLysThr-GluGluGlyLeuGlyPr 533
Db 1608 CAGGTAAAGTGGAGAGCATGT-GACAGAGCAAGAGAA-ACCTGAAGAGGGCTTGGGCC 1665
QY 533 oAsnLeuLysSerLeuValThrMetLeuMetLeuMetLeuMetMetPheAlaValHI 553
Db 1666 CAACATCAAAAGCATTTGACCATGCTGATGCTCATGCTCATGATGTTTCGGGTCCA 1725
QY 553 sCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAs 573
Db 1726 CTGCACGTGGGTCAACACACGCTACTCCAGTCCAGTGTGGTCTTCCCTCTACAA 1785
QY 573 nHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGI 593
Db 1786 TCATGATGGTACCAGGAATATTAGATGATTTTAGAGAACGCTACTTTTGGCTGAGACA 1845
QY 593 nAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGI 613
Db 1846 AAACACGATGATGACACGCGCGGTCTGTGTGGAGTACGGCTATCAGATTGCTGG 1905
QY 613 yMetAlaSerArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisleAlaLe 633
Db 1906 CATGCCCAACAGGACCACTCTGGTGGATAACACACCTGGAAACACAGCCACATCCCACT 1965
QY 633 uValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr-LeuA 653
Db 1966 GGTCCGAAAGATGTTCTTCCATGAAACGGCCGCTATAAAATCATGAGTTC-CTTGG 2024
QY 653 sValAspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleA 673
Db 2025 ATGTCGATTTATGTTGGTTATTTTCGAGGAGTGTGCTATTTCGGGGACGATATCA 2084
QY 673 snLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluS 693
Db 2085 ACAGTTCTCTGGATGTGAGGATAGCTGAAAGGGAGCATCCCAAGACATCCGGGAAG 2144
QY 693 ex-AspTyrPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThrLe 712
Db 2145 G-TGACTATTTTAC-CCAGCAGGAGAGTTCGAGTAGACAAAGCTGGTCTCTACTCT 2202
QY 712 uLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPh 732
Db 2203 GTTAAACTGCCTTATGATATAAAATGTCATACAGATTTGGAGAAATGAGCTAGATT 2262
QY 732 eaArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleY 752
Db 2263 TGCACCTCCCCAGGCTTTGACCAACACGTAATGCTGAGATTGGAAATAAAGACATTAA 2322
QY 752 sPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysVa 772
Db 2323 ATTCAGCATTTGGAGGAGCTTTTACATCAGACACTGCTGTGTCAGGATATATAAGT 2382
QY 772 lLysAlaProAspAsnArgGluThrLeu 781
Db 2383 GAAAGCACCTGACAAACAGGGAGACACTA 2410

RESULT 3

US-09-945-527-62

; Sequence 62, Application US/09945527
; Publication No. US20030055588A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20030055588A1el Nucleic Acid Molecules Encoding
; FILE REFERENCE: 35800/237985
; CURRENT APPLICATION NUMBER: US/09/945,527
; CURRENT FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 1828
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1767, 1771
; OTHER INFORMATION: n = A,T,C or G
US-09-945-527-62

Alignment Scores:
Pred. No.: 0 Length: 1828
Score: 355.00 Matches: 355
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.98% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x US-09-945-527-62 (1-1828)

QY 472 SerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGlyAspMetLysArgGlu 491
Db 70 TCTGCATATGCCCTTTTCAATGTTTGGACCATATTGGGGATGACATGAAAGGAA 129
QY 492 AsnProProValGluAspSerSerAspGluAspAspLysArgAsnGlnGlyAsnLeuTyr 511
Db 130 AATCCACCTGTGGAGGACAGCATGATGAGATGACAAAGAAACCAAGAAATTTGTAT 189
QY 512 AspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLysThrGluGluGlyLeu 531
Db 190 GATTAAGCAGGTAAAGTGAAGAAACATGCACTGAACAGGAAACAACTGAAGAGGATTA 249
QY 532 GlyProAsnLeuLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAla 551
Db 250 GGCCTTATATAAAAGCATTTGCACCAATGCTACTCTAGTCCAGTGTAGTCTGGCTCA 369
QY 552 ValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSer 571
Db 310 GTCCACTGTACCTGGGTCAACAACATGCTACTCTAGTCCAGTGTAGTCTGGCTCA 369
QY 572 TyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeu 591
Db 370 TACAATCATGTGGCACCAGGAATATCTTAGATGATTTTAGAGAGCTTTACTTTTGGCTA 429
QY 592 ArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIle 611
Db 430 AGGCAAAATACAGATGACATGACCGAGTAATGCTTGTGGGATTTAGCTATCAGATA 489
QY 612 AlaGlyMetAlaSerArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 631
Db 490 GCTGGAATGGCTAATAGAACTAGCTTGGTGGATAATAAACACCTGGAAATTAACAGCCACATA 549
QY 632 AlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr 651
Db 550 GCATCTGTGGGAAAGACTATGCTTTCTAATGAACAGCAGCCCTAATAATCATGAGGACT 609
QY 652 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 671
Db 610 CTAGATGTAGATTATGTTTGGTTTATTTTGGAGGGGTTTATTGGCTATTCTGCTGATGAT 669
QY 672 IleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArg 691

Db 1981 TTGGAGGGGTATTGGCTATTCTGGTGATATCAACAAATTTCTCTGGATGGTAGG 2040
 Qy 681 lIeAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
 Db 2041 ATAGCTGAAGAGAGACATCCCAAGACATTCGGGAAGTGAATTTTACCCCAACAGGA 2100
 Qy 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
 Db 2101 GAATTCCTGTAGACAAAGCAGGATCCCTACTTTGTGAATTCCTTATGTATATAAATG 2160
 Qy 721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArg 740
 Db 2161 TCATCTACACATTTGGGAATATCAGCTGGATTTTCGTACACCCCGAGTTTGACCGA 2220
 Qy 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPhe 760
 Db 2221 ACACGTAATGCTGAGATTGGAATTAAGGACATTTAAATTCAAACATTTGGAAGAAGCCTTT 2280
 Qy 761 ThrSerGluHisTyrLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
 Db 2281 ACATCAGACACTGGCTGTGTAGATATATAGATATAAGATATAAGCCTGNTAACAGGAGACA 2340
 Qy 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
 Db 2341 TTAGATCAACAACCTCGAGTCACCAACATTTTCCCAACAGAGATATTGTCAAGAAG 2400
 Qy 801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
 Db 2401 ACTACCAAGAAGAGCGTGGCTACATTAATAAATPAGCTGTTTTTAAGAAGAAGCAAGAA 2460
 Qy 821 lIeSerLysLysThrVal 826
 Db 2461 ATATCTAAGAAGACTGTT 2478

RESULT 2

US-10-028-384-3
 ; Sequence 3, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREAU, Claude
 ; APPLICANT: MCBRIDE, Kevin
 ; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
 ; FILE REFERENCE: 5600-74
 ; CURRENT APPLICATION NUMBER: US/10/028,384
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 2710
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: AK018758
 ; DATABASE ENTRY DATE: 2001-07-05
 ; RELEVANT RESIDUES: (1)..(2469)
 US-10-028-384-3

Alignment Scores:

Pred. No.: 0 Length: 2710
 Score: 365.00 Matches: 720
 Percent Similarity: 98.09% Conservatives: 0
 Best Local Similarity: 98.09% Mismatches: 7
 Query Match: 44.19% Indels: 14
 DB: 15 Gaps: 0

US-10-028-384-2 (1-826) x US-10-028-384-3 (1-2710)

Qy 55 ProAlaGlyLeuSerGlyLeuSerGlnProAlaGlyTrpGlnSerLeuSerPhe 74
 Db 230 CCCGGGGGGCTGTCCGGGGGCTTGTGCGAGCGCGGGTGGCGAGTGTCTCTCTTC 289

Qy 75 ThrIleLeuPheLeuAlaTrpIleAlaGlyPheSerSerArgLeuPheAlaValIleArg 94
 Db 290 ACCATCTCTTCTCGGCTGGCTGGCGGCTTCAGCTCGGCTCTTCGCGCTCATCCGC 349
 Qy 95 PheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgSerThrHisLeu 114
 Db 350 TTGAGAGCATCATCCAGAGTTCGACCGTGGTTTAACTATAGATCAACATCATCTT 409
 Qy 115 AlaSerHisGlyPheTyrGluPheLeuAsnTyrPheAspGluArgAlaTyrProLeu 134
 Db 410 GCATCTCATGGATCTATGAGTTTCTAAATTTGGTTTGAATGAAGAGCATGTATCCCATG 469
 Qy 135 GlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHis 154
 Db 470 GGAAGATAGTGGGTGGCGCCCTTACCCAGGGTGTGATGAATACAGCTGGCTTATTCAT 529
 Qy 155 TrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspValCysValPheLeuAla 174
 Db 530 TGGATTTTAAATACATTAACATTAACAGTTTCACATAAGAGATGTGTGTATTCCTTGA 589
 Qy 175 ProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTyrAsn 194
 Db 590 CCAACTTTTAGCGGCTTACATCCATCTAGTTCCTGCTTAACTAGAGAACTGTGGAA 649
 Qy 195 GlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValProGlyTyrIleSerArg 214
 Db 650 CAAGGAGCAGGACTTCTAGCTGCTTCAATGCTATCGTACCAAGGATACATATCTCG 709
 Qy 215 SerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 234
 Db 710 TCAGTGGCGGATCCTTTGATTAATGAGGCATTTGCCATTTTGGCTTCAGTTACATTAC 769
 Qy 235 TyrIleuTrpValLysSerValLysThrGlySerValPheThrPheMetCysCysLeu 254
 Db 770 TACTTATGGTAAAGTCTGTGAAGACCGGCTGTGTCTCGACAATGTGTGTGTCTGTTG 829
 Qy 255 SerTyrPheTyrMetValSerAlaTyrGlyGlyTyrValPheIleIleAsnLeuIlePro 274
 Db 830 TCATATTTCTACATGGTCTCTGGTGGGAGGATTAATGTTTCATCATCAACTATCCCT 889
 Qy 275 LeuHisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyr 294
 Db 890 CTCATGTGTTTGT 949
 Qy 295 SerThrPhe-TyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPheG 314
 Db 950 AGCAC-TTTGTACATTTGGGTTTAAATATCCATGCGAGATACCTTTTGGGATTTCA 1008
 Qy 314 nProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTy 334
 Db 1009 GCCAATCAGAAACAAGCGAGCACATGGCAGCTGCTGTCTTTGGCTGTCTGCAAGCTTA 1068
 Qy 334 rAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLysPhePh 354
 Db 1069 CGCTTTTTCAGTATCTGAGAGACCGGTTGAAAACAGAGATTCAGACCCCTTTCTT 1128
 Qy 354 eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyTh 374
 Db 1129 TTTGGGTGTCTCACTAGCTGCGAGCGCTGTGTCTCTCTAGTGTCTATCTGACATACAC 1188
 Qy 374 rGlyTyrIleAlaProTyrSerGlyArgPheTyrSerLeuTyrAspThrGlyTyrAlaTy 394
 Db 1189 AGGTATATTGCAACATGGAGTGGCAGGTTTATTTCACTATGGGATATCTGGGTATGCAA 1248
 Qy 394 sIleHisIleProIleAlaSerValSerGluHisGlnProThrThrTrpValSerPh 414
 Db 1249 AATACACATTCCTCAATTAATTCATCACTGCTGCTGAACATCAGCCTACGATGGGTGTCTTT 1308
 Qy 414 ePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTyrPheCysIleLy 434
 Db 1309 CTTCCTTGTATCTACATATCTTGTATGTACCTTCCAGCAGGCGCTATGGTCTGCACTCA 1368
 Qy 434 sAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAl 454

; OTHER INFORMATION:

US-10-028-384-1

Alignment Scores:

Pred. No.: 0 Length: 2481
 Score: 826.00 Matches: 826
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 15 Gaps: 0

US-10-028-384-2 (1-826) x US-10-028-384-1 (1-2481)

QY 1 MetAlaGluProSerAlaProGluSerIlyshisysSerSerLeuAsnSerSerProTrp 20
 Db 1 ATGGCGAGCCCTCGGCCCGGAGAGACAGCAAGTCGTCCTCAACTCGTCCCGGTG 60
 QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 Db 61 AGTGGCCCTCATGGCCCTGGGAAACAGCGCGCACGCCACCGCGCGCGCGCGCGCG 120
 QY 41 AlaHisIysAlaAlaGlyGlyAlaAlaProProIlySerProAlaProAlaGlyLeuSerGly 60
 Db 121 GGGCAACAGCG 180
 QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 Db 181 GGGCTGTGCGACGGCGGTGGTGGCGAGTCGCTCTCTCTCTCCATCTCTCTCTCT 240
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100
 Db 241 TGGCTGTGGCGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAAAGCATCATCCAC 300
 QY 101 GluPheAspProTrpPheAsnTrpArgSerThrHisLeuAlaSerHisGlyPheTrp 120
 Db 301 GAGTTTCGACCCGTGTTAACTATAGATCAACACATCATCTTCATCTCATGGTCTCTAT 360
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGlyGly 140
 Db 361 GAATTTTAAATGGTTTGTATGAAGAGCATGGTATCCATCCTAGGAAGATAGTAGTGGT 420
 QY 141 ThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisIleStrIleLeuAsnThrLeu 160
 Db 421 ACTGTTTACCAGGGTGTATGATATAACCGCTGGCTTATTCATTGGATTTTAAATACATTG 480
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 Db 481 AACATAACTGTTTCATAGAGAGCATGTGTGTCTCTTGCACCACTTTTAGCGGCGCTT 540
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
 Db 541 ACATCTATATCTACTTTCTCTCTTACAAGAGAACTTTTGAACCAAGGAGCAGGACTTTTA 600
 QY 201 AlaAlaCysPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySerPhe 220
 Db 601 GCTGTGTGTTTATTGCTATTGTACAGGCTACATATCTCGGTACAGTAGCTGATCTCTT 660
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpValIlySer 240
 Db 661 GATAATGAAGGCATTGCTATTATTTTGCACCTTCAGTTCACATATTTATGGTAAATCT 720
 QY 241 ValIlySerGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMetVal 260
 Db 721 GTAAATGCTGGGTGAGTCTTTTGGACAAATGTCGTGCTATCTCTATTTCTATATGGTC 780
 QY 261 SerAlaTrpGlyGlyTrpValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
 Db 781 TCTGCTGGGTGGTGTATGTTATTTATCATCAATCTTATCCATGCTGATGTTTGGTGTG 840
 QY 281 LeuLeuMetGlnArgTrpSerIysArgValTrpIleAlaTrpSerThrPheTrpIleVal 300
 Db 841 TTACTGATGAGAGATACAGCAAGAGTCTTACATAGATATAGCACTTTCTACATGTTG 900

QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 Db 901 GGTTTAATATATCAATGTCAGATACCTTTTGTGGATTCCAGCAATCAGAAAGTGAA 960
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeu 340
 Db 961 CACATGGCAGTGGAGGTGCTTTGCAATGCTGCTCAAGCTATGCTTTCTTGAGTATCTG 1020
 QY 341 ArgAspArgLeuThrIysGlnIupheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 Db 1021 AGAGACCGATTAACAAAAACAAGAGTTCAGACCCCTTTCTTTTGGGTGTATCACTAGCT 1080
 QY 361 AlaglyAlaValPheLeuSerValIleTrpLeuThrTrpTrpClyTrpIleAlaProTrp 380
 Db 1081 CGAGTCTGTGTTCTCTTAGTGTCACTATTGACTTATACAGGTACATTCACCATGG 1140
 QY 381 SerGlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaIlyIleHisIleProIleIle 400
 Db 1141 AGTGGCAGGTTTATTCTTCTTGGGATCTGGGTATGCAAAAAATACACATTCCAATTAT 1200
 QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePhePheAspLeuHisIle 420
 Db 1201 GCATCAGTGTGAGCATCAACCTACGACTTGGGTGCTTTCTCTTTGATCTACATAT 1260
 QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleIlyAsnIleAsnAspGluArg 440
 Db 1261 CTTGTATGTACTTCCCGCAGCGCTTGGTTCGTCATCAAAAAATATCAACGATGAAAGA 1320
 QY 441 ValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetValArgLeu 460
 Db 1321 GTATTTGTTGCTCATATGCAATCAGTGTGCTGCTCTTTGCTGGAGTGTATGGTGACCTG 1380
 QY 461 MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
 Db 1381 ATGTTGACTTTGACTCCAGTCGTGTATGCTGCTGCTGCAATTCCTTTTCAATGTTTTT 1440
 QY 481 GluHisTrpLeuGlyAspAspMetIysArgGluAsnProProValGluAspSerSerAsp 500
 Db 1441 GAGCAGCTATTGTTGGGGATGACNTGAAAAGGAAAAATCCACCTGTGGAGCAGCAGTGAT 1500
 QY 501 GluAspAspIysArgAsnGlnGlyAsnLeuTrpAspIysAlaGlyIysValArgIysHis 520
 Db 1501 GAGGATGACAAAAAGAAACCAAGGAAATTTGTATGATAAGCGAGGTAAAGTGAGGAAACAT 1560
 QY 521 AlaThrGluGlnGlyThrGluGluGlyLeuGlyProAsnIleIysSerIleValThr 540
 Db 1561 GCAACTGACAGGAAAAAACTGAAGAGGATTAGGCCCTTAATATAAAAGCAATTTGCACC 1620
 QY 541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
 Db 1621 ATGTTGATGCTGATGCTATTGATGATGTTTGTCTGCTCCACTGTACCTGGGTCAACAGCAAT 1680
 QY 561 AlaTrpSerSerProSerValValLeuAlaSerTrpAsnHisAspGlyThrArgAsnIle 580
 Db 1681 GCCTACTCTAGTCCAGGTGAGTCTGCGCTCTATCAATCATGATGAGCAGGCAATATC 1740
 QY 581 LeuAspAspPheArgGluAlaTrpPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
 Db 1741 TTAGATGATTTAGAGAGCTTACTTTTGGCTTAAGGCAAAATACACATGAAATGACCA 1800
 QY 601 ValMetSerTrpTrpAspTrpGlyTrpGlnIleAlaGlyMetAlaAsnArgThrLeu 620
 Db 1801 GTATGCTCTGGTGGGATATGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG 1860
 QY 621 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyIysAlaMetSerSer 640
 Db 1861 GTGGATAATAACACCTGGAATTAACAGCACATAGCACTGGTGGGAAAGCTATGTCTCT 1920
 QY 641 AsnGlnThrAlaAlaTrpIlyIleMetArgThrLeuAspValAspTrpValLeuValIle 660
 Db 1921 AATGAACAGGAGGCTATAAAATCATGAGGACTTAGATGTAGATTATGTTTGGTTAT 1980
 QY 661 PheGlyGlyValIleGlyTrpSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 23:35:31 ; Search time 1039.95 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 826

Sequence: 1 MAEPSAPESKHKSLSNPW.....GYIKNLVFKGKKSKTKTV 826

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8161859

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-Q/cgn2_1/USPTO.spool/US10028384/runat_14122004_131648_6892/app_query.fasta.1.3740
-DB=PublishedApplications NA -QW=fastap -SUFFIX=Oligo.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0
-NAXLEN=2000000000 -USER=US10028384@cgn 1 1 2124 @runat 14122004_131648_6892
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	826	100.0	2481	15	US-10-028-384-1
2	365	44.2	2710	15	US-10-028-384-3
3	355	43.0	1828	10	US-09-945-527-62
4	351	42.5	1209	15	US-10-106-698-330
5	351	42.5	1209	16	US-10-264-237-412
6	351	42.5	1209	10	US-09-974-879-133
7	351	42.5	1543	10	US-09-305-736-133
8	351	42.5	1543	10	US-09-818-683-133
9	351	42.5	1543	11	US-09-818-683-133
10	351	42.5	1543	16	US-10-621-401-133
11	217	26.3	1114	16	US-10-296-115-629
12	196	23.7	2660	16	US-10-264-049-630
13	166	20.1	500	9	US-09-998-598-1643
14	143	17.3	430	9	US-09-878-178-717
15	143	17.3	430	13	US-10-046-935-717
16	143	17.3	430	14	US-10-146-502-717
17	133	16.1	588	14	US-10-052-283-433
18	119	14.4	387	16	US-10-276-774-173
19	117	14.2	616	9	US-09-879-536-332
20	108	13.1	485	10	US-09-918-995-11283
21	104	12.6	743	10	US-09-945-527-63
22	68	8.2	637	15	US-10-002-631C-183
23	68	8.2	640	15	US-10-002-631C-179
24	54	6.5	474	14	US-10-052-283-454
25	51	6.2	393	16	US-10-276-774-323
26	44	5.3	312	9	US-09-796-692-8870
27	44	5.3	312	14	US-10-040-862-8870
28	44	5.3	312	16	US-10-057-475B-8870
29	44	5.3	312	16	US-10-154-884B-8870
30	44	5.3	312	17	US-10-764-324-8870
31	42	5.1	2417	15	US-10-028-384-7
32	40	4.8	352	9	US-09-878-178-1032
33	40	4.8	352	13	US-10-046-935-1032
34	40	4.8	352	14	US-10-146-502-1032
35	34	4.1	483	10	US-09-918-995-23896
36	33	4.0	1848	15	US-10-128-714-2139
37	33	4.0	1969	15	US-10-128-714-1139
38	33	4.0	2332	15	US-10-128-714-7139
39	33	4.0	2603	15	US-10-128-714-6139
40	33	4.0	3969	15	US-10-128-714-139
41	33	4.0	4603	15	US-10-128-714-5139
42	21	2.5	154	10	US-09-991-936-450
43	20	2.4	502	17	US-10-767-701-6513
44	19	2.3	2256	15	US-10-032-585-6323
45	17	2.1	887	16	US-10-424-599-77697

ALIGNMENTS

RESULT 1
US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPARTIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)

APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 3090
LENGTH: 594
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-3090

Alignment Scores:
Pred. No.: 4,428-07 Length: 594
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.94% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-248-796A-3090 (1-594)

Qy 92 VallleArgpHeGluSerllelleHleHisGluPheAspProTrrpPheAsn 107
Db 151 GTGATTCGATTGGAAGTATTATTTCATGATTCGATTCCTGGTTCAAT 198

Search completed: December 14, 2004, 23:43:35
Job time : 170.937 secs

Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 16600
LENGTH: 560
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-16600

Alignment Scores:
Pred. No.: 1.57e-18 Length: 560
Score: 27.00 Matches: 27
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.27% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-16600 (1-560)

QY 445 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 464
Db 423 CTGTAGCCATCAGTCGGTTTACTTCGCTGGTGATGCGGTTTGAATGACCCCTC 364

QY 465 ThrProValValCysMetLeu 471
Db 363 ACGCCGCTGGTGCGATGCTG 343

RESULT 12
US-09-270-767-28082
Sequence 28082, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 28082
LENGTH: 133
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-28082

Alignment Scores:
Pred. No.: 7.34e-12 Length: 133
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.42% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-28082 (1-133)

QY 305 SerMetGlnIleProPheValGlyPheIleArgThrSerGluHisMetAlaAla 324
Db 68 TCCATGCAGATCCCTTCGTGGGATTCACCGATACCGACCATGACATGGCTGGC 127

RESULT 13
US-09-248-796A-3089
Sequence 3089, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 3089
LENGTH: 867
TYPE: DNA
ORGANISM: Candida albicans
FEATURE:
NAME/KEY: unsure
LOCATION: (25)
OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-248-796A-3089

Alignment Scores:
Pred. No.: 4.87e-10 Length: 867
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.30% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-248-796A-3089 (1-867)

QY 553 HisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSer 571
Db 94 CATGTACTTGGGTACATCATGATGCTTATTCACCATCAGTCTGTTTAGCATCC 150

RESULT 14
US-09-614-221A-318
Sequence 318, Application US/09614221A
Patent No. 6723837
GENERAL INFORMATION:
APPLICANT: Karunanandaa, Balasulojini
APPLICANT: Yu, Jaehyuk
APPLICANT: Kishore, Ganesh M.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH STEROL SYNTHESIS AND METABOLISM
FILE REFERENCE: 16516.075
CURRENT APPLICATION NUMBER: US/09/614,221A
CURRENT FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/142,981
PRIOR FILING DATE: 1999-07-12
NUMBER OF SEQ ID NOS: 626
SEQ ID NO 318
LENGTH: 2157
TYPE: DNA
ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-318

Alignment Scores:
Pred. No.: 1.4e-07 Length: 2157
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.06% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-614-221A-318 (1-2157)

QY 392 TyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnPro 408
Db 1009 TACGCAAGATCCACATTCCTATCATTCCTCCGTTCCGAACATCAACCC 1059

RESULT 15
US-09-248-796A-3090
Sequence 3090, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:

```
US-09-270-767-158/c
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

Alignment Scores:
Pred. No.: 1,53e-30 Length: 1660
Score: 39.00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.72% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-158 (1-1660)
QY 652 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 671
Db 673 CTTCAGCTGGACTAGCTTTGGTGTATCTTGGCGGTGTGATCGGCTATTCGGCGATGAT 614
QY 672 IleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIle 690
Db 613 ATCAACAAGTTCTCTGGATGTCGGAATTCGGAGGGAGAGCATCCCAAGGACATT 557

RESULT 8
US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Alignment Scores:
Pred. No.: 1,53e-30 Length: 1660
Score: 39.00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.72% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-15440 (1-1660)
QY 652 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 671
Db 673 CTTCAGCTGGACTAGCTTTGGTGTATCTTGGCGGTGTGATCGGCTATTCGGCGATGAT 614
QY 672 IleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIle 690
Db 613 ATCAACAAGTTCTCTGGATGTCGGAATTCGGAGGGAGAGCATCCCAAGGACATT 557

RESULT 9
US-09-270-767-12331
; Sequence 12331, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12331
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-12331

Alignment Scores:
Pred. No.: 1,9e-21 Length: 900
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.63% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-12331 (1-900)
QY 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIle 99
Db 160 GCCTGGCTGGCGGATTTCTCTCGCTCTCGCGTCATCGTTTCGATCGATTATC 219
QY 100 HisGluPheAspProTyrPheAsnTyrArg 109
Db 220 CATGAGTTTGATCGTGGTTCACACTACCGG 249

RESULT 10
US-09-270-767-1318/c
; Sequence 1318, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1318
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1318

Alignment Scores:
Pred. No.: 1,57e-18 Length: 560
Score: 27.00 Matches: 27
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.27% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-1318 (1-560)
QY 445 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 464
Db 423 CTGTACGCCATCACTCGGTTTACTTCGCTGGTGTGATGCGTTTGATGTGACCCCTC 364
QY 465 ThrProValValCysMetLeu 471
Db 363 ACGCGGTGGTGTGCATGCTG 343

RESULT 11
US-09-270-767-16600/c
; Sequence 16600, Application US/09270767
```

QY 285 TyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSer 305
Db 63 TACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTGTGGGTTTAATATTATCA 122
QY 306 MetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAla 325
Db 123 ATGCAGATACCTTTTGTGGGATTCAGCCAAATCAGAACAGTGAACACATGGCAGTCCA 182
QY 326 GlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThr 345
Db 183 GGTGCTTTTGCAATGCTCCTCAAGCTATGCTTTCTTTCAGTATCTGAGAGCCGATTACA 242
QY 346 LysGln 347
Db 243 AAACAA 248

RESULT 4

US-09-513-999C-21090
; Sequence 21090, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Malne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 21090
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-21090

Alignment Scores:
Pred. No.: 7,586-72 Length: 245
Score: 78.00 Matches: 78
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.44% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-513-999C-21090 (1-245)

QY 185 ThrPheLeuThrArgGluLeuTyrAsnGlnGlyAlaGlyLeuLeuAlaCysPhe 204
Db 3 ACTTTCCTGCTTACAGAGACTTTGGAACCAAGAGCAGGACTTTTAGCTGCTGTTT 62
QY 205 IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspGluGly 224
Db 63 ATTGCTATTGACCAAGGCTACATATCTCGTTCAGTAGCTGATCCCTTGTATGAGGC 122
QY 225 IleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValLysSerValLysThrGly 244
Db 123 ATTGCTATTTTGGACITTCAGTTACATATCTTTATGGGTAAATCTGTAAACCTGG 182
QY 245 SerValPheThrThrMetCysCysCysLeuSerTyrPheTyrMetValSerAla 262
Db 183 TCAGTTTTTGGCAATGTGCTGCTTATCTCTTATTTCTATATGCTCTCTGCT 236

RESULT 5

US-09-270-767-158
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

Alignment Scores:
Pred. No.: 1,178-33 Length: 1660
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.08% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-158 (1-1660)

QY 599 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 618
Db 854 GCTCGCTTATGCTTTGGTGGATTACGATACCGGTAATGCGCAACAGAACG 913
QY 619 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 638
Db 914 ACGTAGTGGATATATACGTGGACATATAGTCATAGCGCTGTTGGCAAGCAATG 973
QY 639 SerSer 640
Db 974 TCTTCA 979

RESULT 6

US-09-270-767-15440
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Alignment Scores:
Pred. No.: 1,178-33 Length: 1660
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.08% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-15440 (1-1660)

QY 599 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 618
Db 854 GCTCGCTTATGCTTTGGTGGATTACGATACCGGTAATGCGCAACAGAACG 913
QY 619 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 638
Db 914 ACGTAGTGGATATATACGTGGACATATAGTCATAGCGCTGTTGGCAAGCAATG 973
QY 639 SerSer 640
Db 974 TCTTCA 979

RESULT 7

QY 562 TyrSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeu 581
Db 2 TACTCTAGTCCAAAGTGTAGTCTCGCTCATACATCATGATGCACACAGGAATCTTA 61
QY 582 AspAspPheArgGluAlaValTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgVal 601
Db 62 GATGATTTTAGAAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACGAGTA 121
QY 602 MetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuVal 621
Db 122 ATGCTCTGGTGATTAATGCTATCAGATAGCTGGATGGCTAATAGAACTACGTTGGTG 181
QY 622 AspaenAnthrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsn 641
Db 182 GATAATAACACCTGGAATTAACAGCCACATAGCACTGGTGGGAAAGCTATGCTCTCTAAT 241
QY 642 GluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValIlePhe 661
Db 242 GAAACAGCAGCCTATAAATCATGAGGACTCTAGATGATGATTTTGGTTATTTT 301
QY 662 GlyGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrpMetValArgIle 681
Db 302 GGAGGGGTATTTGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGGATA 361
QY 682 AlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGlu 701
Db 362 GCTGAAGGAGACATCCCAAGACATTCGGGAAAGTGCATATTTTACCCACAGGGAGAA 421
QY 702 PheArgValAspLysAlaGlySerProThrIleuLeuAsnCysLeuMetTyrLysMetSer 721
Db 422 TTCGCTGTAGACAAAGCAGGATCCCTACTTTGTTGAATTCGCTTAATGATATAAATGCTCA 481
QY 722 TyrTyrArgPheGlyGluMetGln 729
Db 482 TACTACAGTTTGGAGAATGCAG 505

RESULT 2

US-09-328-111-332
; Sequence 332, Application US/09328111
; Patent No. 6262333

GENERAL INFORMATION:

; APPLICANT: Endeigo, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Detti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 332
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(616)
; OTHER INFORMATION: n = A,T,C or G

US-09-328-111-332

Alignment Scores:

Pred. No.: 5,73e-112 Length: 616
Score: 117.00 Matches: 117
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.16% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-2 (1-826) x US-09-328-111-332 (1-616)

QY 208 ValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIle 227
Db 5 GTACCAGGCTACATATCTCGTCACTAGTCGATCTTTGATAATGAAGGCATTCCTATT 64
QY 228 PheAlaLeuGlnPheThrTyrTyrLeuTrpValIysSerValIysThrGlySerValPhe 247
Db 65 TTTCGACTTCAGTTTCACTACTATTATTATGGGAAAAATCTGTAAAAACCTGGGTCTATTTT 124
QY 248 TrpThrMetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrVal 267
Db 125 TGGACAAATGCTGCTCTATCTTATCTTATATATGCTCTCTGCTTGGGTGGTTATGTA 184
QY 268 PheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSer 287
Db 185 TTATCATCAATCTTATTCCTCACTGTCATGATTTTGTGTTTACTGATCGAGATACAGC 244
QY 288 LysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGln 307
Db 245 AAAAGAGTCTACATAGCATATAGCACTTTTACATTGTGGGTTTAATATTATCATATGCAG 304
QY 308 IleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAla 324
Db 305 ATACCTTTTGTGGGATTCAGCAATCAAGCAAGTGAACACATGGCAGCT 355

RESULT 3

US-09-513-999C-1438
; Sequence 1438, Application US/09513999C
; Patent No. 6783961

GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59,US2,REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1438
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 54..248
; US-09-513-999C-1438

Alignment Scores:

Pred. No.: 5,41e-76 Length: 250
Score: 82.00 Matches: 82
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.93% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-513-999C-1438 (1-250)

QY 266 TyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArg 285
Db 3 TAATGTTTATCATCAATCTTATTCACATGATGTTTGTGTTTACTGATGCAGAGA 62

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 160.937 Seconds
(without alignments)
3648.073 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 826

Sequence: 1 MAEPSAPSKHKSLSNSPW.....GYKNKLVPKKGKISKKTU 826

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n model -DEV=xlh
-Q/cgn2_1/USPTC_spool/US10028384/runat_14122004_131646_6916/app_query.fasta.1.3740
-DB=Issued Patents NA -QFWT=fastcap -SUFFIX=Oligo.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human0.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN 1 1 259 @runat_14122004_131646_6816 -NCPUS=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -HEADS=1 -XGAPOP=60 -XGAEXT=60 -FGAPOP=6
-FGAEXT=7 -YGAPOP=60 -YGAEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	168	20.3	507	US-09-513-999C-1965	Sequence 1965, Ap
2	117	14.2	616	US-09-328-111-332	Sequence 332, App
3	82	9.9	250	US-09-513-999C-1438	Sequence 1438, Ap
4	78	9.4	245	US-09-513-999C-21090	Sequence 21090, A
5	42	5.1	1660	US-09-270-767-158	Sequence 158, App
6	42	5.1	1660	US-09-270-767-15440	Sequence 15440, App
C 7	39	4.7	1660	US-09-270-767-158	Sequence 158, App
C 8	39	4.7	1660	US-09-270-767-15440	Sequence 15440, A
C 9	30	3.6	900	US-09-270-767-12331	Sequence 12331, A
C 10	27	3.3	560	US-09-270-767-1318	Sequence 1318, Ap
C 11	27	3.3	560	US-09-270-767-16600	Sequence 16600, A
12	20	2.4	133	US-09-270-767-28082	Sequence 28082, A

13	19	2.3	867	4	US-09-248-796A-3089	Sequence 3089, Ap
14	17	2.1	2157	4	US-09-614-221A-318	Sequence 318, App
15	16	1.9	594	4	US-09-248-796A-3090	Sequence 3090, Ap
16	15	1.8	487	3	US-09-385-982-213	Sequence 213, App
17	14	1.7	302	4	US-09-313-294A-6869	Sequence 6869, Ap
18	13	1.6	914	4	US-09-270-767-12856	Sequence 12856, A
19	13	1.6	1386	4	US-09-270-767-11648	Sequence 11648, A
20	12	1.5	268	4	US-09-313-294A-4834	Sequence 4834, Ap
21	12	1.5	291	4	US-09-313-294A-4834	Sequence 4834, Ap
22	12	1.5	299	4	US-09-313-294A-1303	Sequence 1303, Ap
23	11	1.3	307	4	US-09-702-705-375	Sequence 375, App
24	11	1.3	307	4	US-09-702-705-588	Sequence 588, App
25	11	1.3	307	4	US-09-702-705-1271	Sequence 1271, Ap
26	11	1.3	307	4	US-09-736-457-375	Sequence 375, App
27	11	1.3	307	4	US-09-736-457-588	Sequence 588, App
28	11	1.3	307	4	US-09-736-457-1271	Sequence 1271, Ap
29	11	1.3	307	4	US-09-614-124B-375	Sequence 375, App
30	11	1.3	307	4	US-09-614-124B-588	Sequence 588, App
31	11	1.3	307	4	US-09-614-124B-1271	Sequence 1271, Ap
32	11	1.3	307	4	US-09-671-325-375	Sequence 375, App
33	11	1.3	307	4	US-09-671-325-588	Sequence 588, App
34	11	1.3	307	4	US-09-671-325-1271	Sequence 1271, Ap
35	11	1.3	307	4	US-09-589-184-375	Sequence 375, App
36	11	1.3	307	4	US-09-589-184-588	Sequence 588, App
37	11	1.3	307	4	US-09-589-824-375	Sequence 375, App
38	11	1.3	307	4	US-09-589-824-588	Sequence 588, App
39	11	1.3	307	4	US-09-589-824-1271	Sequence 1271, Ap
40	11	1.3	487	4	US-09-702-705-1655	Sequence 1655, Ap
41	11	1.3	487	4	US-09-736-457-1655	Sequence 1655, Ap
42	11	1.3	487	4	US-09-614-124B-1655	Sequence 1655, Ap
43	11	1.3	487	4	US-09-671-325-1655	Sequence 1655, Ap
44	11	1.3	487	4	US-09-658-824-1655	Sequence 1655, Ap
45	10	1.2	109	4	US-09-270-767-28703	Sequence 28703, A

ALIGNMENTS

RESULT 1
US-09-513-999C-1965
; Sequence 1965, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Guclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513, 999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1965
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 122..505
US-09-513-999C-1965

Alignment Scores:
Pred No.: 5,08e-165 Length: 507
Score: 168.00 Matches: 168
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.34% Indels: 0
DB: 4 Gaps: 0
US-10-028-384-2 (1-826) x US-09-513-999C-1965 (1-507)

QY	761	ThrSerGluHisTrpLeuValArgIleTyrLysValIysAlaProAspAsnArgGluThr	780
Db	663	ACATCAGACACTGGCTTGTAGCATATATAAGTAAAGCACCTCATACAGGGGAC	722
QY	781	LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys	800
Db	723	TTAGATCACAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGTATTTGTCAAGAAG	782
QY	801	ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys	820
Db	783	ACTACCAAAAGAGCGTGGCTACATTAATAATAGCTGCTTTTAAAGAGGCAAGAA	842
QY	821	IleSerLysLysThrVal	826
Db	843	ATATCTAAGAAGACTGTT	860

Search completed: December 14, 2004, 22:36:14
Job time : 9206.03 secs

```

QY 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThr 577
Db 119 ACAAGCAATGCTACTCTAGTCCAAAGTGTAGTCTGCCTCAACATCATGATGCAC 178
QY 578 ArgAsnIleLeuAspAspPheArgGluAlaTyrPheThrLeuArgGlnAsnThrAspGlu 597
Db 179 AGGAATATCTTAGATGATTTTAGAGAGCTTACCTTTGGCTAAGGCAAAATACAGATGA 238
QY 598 HisAlaArgValMetSerThrTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 617
Db 239 CATGCACAGATGATGCTCTGTGGATATGCTGATCAGATAGCTGAATGCGCTAATAGA 298
QY 618 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 637
Db 299 ACTACGTTGGTGGATAATAACACCTGGATAAAGCACACCATAGCACTGGTGGGAAAGCT 358
QY 638 MetSerSerAsnGluThrAlaAlaTyrIleMetArgThrLeuAspValAspTyrVal 657
Db 359 ATGCTCTCTATGAACACAGACGCTATTAATCATGAGGACTCTAGATGATGATTAATGT 418
QY 658 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 677
Db 419 TTGGTATATTTTGGAGGGTATTTGGCTATTTCTGCTGATGATATCAACAAATTTCTCTGG 478
QY 678 MetValArgIleAlaGluGlyGluHisProIleAspIleArgGluSerAspTyrPheThr 697
Db 479 ATGGTTAGATAGCTGAAGAGACATCCCAAGACATTCGGAAAGTGAATTTTACC 538
QY 698 ProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMet 717
Db 539 CCACAGGAGAAATTCCTGTAGACAAAGCAGGATCCCTACTTTTGTGTAATTCCTATG 598
QY 718 TyrIleMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGly 737
Db 599 TATAAATGTCTACTACAGATTTGGAGAAATGCAGCTGGATTTTGTACACCCCGAGT 658
QY 738 PheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGlu 757
Db 659 TTTGCCGAACACGCTAATGCTGAGATTCGAAATGAAGACATTAATTAATCAACATTTGGA 718
QY 758 GluAlaPheThrSerGluHisThrLeuValArgIleTyrLysValLysAlaProAspAsn 777
Db 719 GAAGCTTTATCATCAGAACACTGCTTGTAGATATATTAATTAATTAAGCAACCTGATAAC 778
QY 778 ArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeu 797
Db 779 AGGAGACATTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAAATTTTG 838
QY 798 SerLysLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLys 817
Db 839 TCNAGAGACTACCAAAAGAGCGTGGCTACATTAATAAATAAGCTGGTTTTTAAGAA 898
QY 818 GlyLysLysIleSerLysLysThrVal 826
Db 899 GCGAAGAAATATCTAAGAGAGCTGTT 925

RESULT 15
AX882932
LOCUS 2284 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 17837 from Patent EP1074617.
ACCESSION AX882932
VERSION AX882932.1 GI:40037833
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primers for synthesizing full-length cDNA and their use
Patent: EP 1074617-A 17837 07-FEB-2001;

```

Research Association for Biotechnology (J2P)

FEATURES

Location/Qualifiers

1. 2284

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

3. 863

/notes="unnamed protein product"

/codon_start=1

/protein_id="CAE91783.1"

/db_xref="GI:40037834"

REAFYLRQNTDEHAFVMSWDYQYIAGMANRTLVNDNTWNSHIALVAGKAMSNE
 TAAKIMRLDLDYLVIVFGVIGSGDDINKFLMWRVIAEGEHPKDIRESDYFTPOG
 EFPVDKAGSPTLLNCLMKVRSYRFGEMOLDPRTPPGDRTRNAEIGNKD:KFHLE
 APTSEHLVRIYKVRAPDNRELDHKPRVTNIFPKQKYLSSKTKTKRKGVIKNKLVFK
 KGKISKIVT"

ORIGIN

Alignment Scores:

Pred. NO.: 7,72e-297 Length: 2284
 Score: 286.00 Matches: 286
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 34.62% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-2 (1-826) x AX882932 (1-2284)

```

QY 541 MetLeuMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
Db 3 ATGTGTGATGCTGATGCTATTGATGATGTTGCTGCCACTGTACCTGGGTCAACAGCAAT 62
QY 561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
Db 63 GCCTACTCTAGTCCAGTAGTGTAGTCTGGCCTCATCAATCATGATGAGGACCCAGGAATATC 122
QY 581 LeuAspAspPheArgGluAlaTyrPheThrLeuArgGlnAsnThrAspGluHisAlaArg 600
Db 123 TTAGATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACGA 182
QY 601 ValMetSerThrTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
Db 183 GTATGCTTTGGTGGGATTTATGCTATCAGATAGTGTGGAATGGCTAATAGAACTACGTTG 242
QY 621 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
Db 243 GTGATAATAACACCTTGAATTAACAGCCACATAGCAGCTGGTGGGAAAAGCTATGCTTCT 302
QY 641 AsnGluThrAlaAlaTyrIleMetArgThrLeuAspValAspTyrValLeuValIle 660
Db 303 AATGAACAGCAGCCCTATAAATCATGAGGACTCTAGATGTAGATTATGTTTGGTTATT 362
QY 661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
Db 363 TTTGGAGGGGTATTGCTGCTATTCTGGTGATGATATCAACAAATTTCTCTGGATGTTAGG 422
QY 681 IleAlaGluGlyLysIleProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
Db 423 ATAGCTGAAGGAGAAATCCCAAGACATTCGGGAAAGTGACTATTTTACCCCAAGGGA 482
QY 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
Db 483 GAATTCGCTGTAGACAAAGCAGGATCCCTACTTTGTTGAATTTGCTTATGATATAAATG 542
QY 721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArg 740
Db 543 TCATACTACAGATTTGGAGAAATGACGCTGGGATTTTCGTACACCCCAAGGTTTTCACCGA 602
QY 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPhe 760
Db 603 ACACGTAATGCTGAGATTGGAATAAGGACATTAATAATCAAAACATTTTGAAGAGACCTTT 662

```

ACCESSION AX017997
 VERSION AX017997.1 GI:10042496
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 REFERENCE Schmitt, A., Specht, T., Dahl, E., Hinzmann, B., Rosenthal, A. and
 AUTHORS Pilarsky, C.
 TITLE Human nucleic acid sequences from prostate tissue
 JOURNAL Patent: WO 9946375-A 232 16-SEP-1999;
 SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
 BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUEH GENOMFORSCHUN
 (DE); PILARSKY CHRISTIAN (DE)
 FEATURES
 source Location/Qualifiers
 1..2503
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,93e-300 Length: 2503
 Score: 289.00 Matches: 289
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 34.99% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-2 (1-826) x AX017997 (1-2503)
 QY 538 IleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
 DB 59 ATTGTCACCATGTTGATGCTGATGCTATTGATGATGTTTGGCTCCACGTACCTGGGTC 118
 QY 558 ThrSerAsnAlaTySerSerProSerValValValLeuAlaSerTySerHisAspGlyThr 577
 DB 119 ACAAGCAATGCTACTCTAGTCCCAAGTGTAGTCTGGCTCATCAATCATGATGGCACC 178
 QY 578 ArgAsnIleLeuAspAspPheArgGluAlaTyPheTrpLeuArgGlnAsnThrAspGlu 597
 DB 179 AGGAATATCTTAGATGATGATTTAGAGAAGCTTACTTTGGCTAAGGCAAAATACAGATGAA 238
 QY 598 HisAlaArgValMetSerTrpTrpAspTyArgGlyTyGlnIleAlaGlyMetAlaAsnArg 617
 DB 239 CATGCACGAGTAATCTCTTGGTGGGATATGGCTATCAGATAGCTGGAAATGGCTAATAGA 298
 QY 618 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 637
 DB 299 ACTACGTTGGTGGATATAACACCTGGATAAAGCAAGCAGACATAGCACTGGTGGGAAAGCT 358
 QY 638 MetSerSerAsnGluThrAlaAlaTyLysIleMetArgThrLeuAspValAspTrpVal 657
 DB 359 ATGCTCTCTAAGAAACAGCAGCTTAAATATCATGAGGACTCTAGATGATGATGATGTT 418
 QY 658 LeuValIlePheGlyGlyValIleGlyTySerGlyAspAspIleHisLeuLysPheLeuTrp 677
 DB 419 TTGGTTATTTTGGAGGGGTATTGGCTATTCTGGTATCTGGTATGATCAACAAATTTCTGG 478
 QY 678 MetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTrpPheThr 697
 DB 479 ATGGTTAGATAGCTGAGGAGAACATCCCAAGACATTCGGGAAGTACTATTATTAC 538
 QY 698 ProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMet 717
 DB 539 CCACAGGAGAAATTCCTGTAGACAAAGCAGGATCCCTACTCTTGTGATTTGCTTATG 598
 QY 718 TyLysMetSerTyTrpArgPheGlyGluMetGlnLeuAspPheArgThrProGly 737
 DB 599 TATAAATGCTACTACTACAGATTGGAGAAATGACAGCTGGATTTTGGTACACCCAGGT 658
 QY 738 PheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGlu 757

DB 659 TTTCACCGAACACGTAATGCTGAGATTGGAATAGGACATTAATAATCAACATTTGGAA 718
 QY 758 GluAlaPheThrSerGluHisTrpLeuValArgIleTyLysValLysAlaProAspAsn 777
 DB 719 GAAGCTTTACATCAGACACCTGGCTTGTAGGATATATAAGTAAAGACCTGATAAC 778
 QY 778 ArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyLeu 797
 DB 779 AGGAGACATTAGATCACAACCTCGAGTCACCAACATTTCCCAACACAGAGTATTG 838
 QY 798 SerLysLysThrThrLysArgLysArgGlyTyLysAsnLysLeuValPheLysLys 817
 DB 839 TCAAAGACACTACCAAAAGGAGGCTGGCTACATTAATAAATAGCTGGTTTAAAGAA 898
 QY 818 GlyLysLysIleSerLysLysThrVal 826
 DB 899 GCGAAGAAATATCTAAGAAGACTGTT 925
 RESULT 14
 BD137303 2503 bp DNA linear PAT 18-SEP-2002
 LOCUS Human nucleic acid sequence originating in prostatic tissue.
 DEFINITION BD137303
 ACCESSION BD137303.1 GI:23232248
 VERSION JP 2002505878-A/64
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2503)
 AUTHORS Specht, T., Hinzmann, B., Schmitt, A., Pilarsky, C., Dahl, E. and
 Rosenthal, A.
 TITLE Human nucleic acid sequence originating in prostatic tissue
 JOURNAL Patent: JP 2002505878-A 64 26-FEB-2002;
 METAGEN GESELLSCHAFT FUER GENOM FORSCHUNG MBH
 COMMENT OS Homo sapiens (human)
 EN JP 2002505878-A/64
 PD 26-FEB-2002
 PF 09-MAR-1999 JP 2000535742
 PR 10-MAR-1998 DE 198 11 194.0
 PI THOMAS SPECHT, BERND HINZMANN, ARMIN SCHMITT, CHRISTIAN PILARSKY,
 EDGAR DAHL,
 PI ANDRE ROSENTHAL
 PC C12N15/09, A61K38/00, A61K39/395, A61K48/00, A61P13/08,
 A61P35/00,
 PC C07K14/47, C07K16/18, C12N1/19, C12N1/21, C12N5/10, C12N15/02, PC
 C12P21/02,
 PC C12P21/08, G01N33/53, C12N15/00, A61K37/02, C12N5/00, C12N15/00 CC
 Human nucleic acid sequence originating in prostatic tissue PH
 Key Location/Qualifiers
 FT source 1..2503
 /organism="Homo sapiens (human)"
 FEATURES
 source Location/Qualifiers
 1..2503
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,93e-300 Length: 2503
 Score: 289.00 Matches: 289
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 34.99% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-2 (1-826) x BD137303 (1-2503)
 QY 538 IleValThrMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
 DB 59 ATTGTCACCATGTTGATGCTGATGCTATTGATGATGTTTGGCTCCACGTACCTGGGTC 118

QY 530 TrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTrpGlyTyr 609
 Db 241 TGGCTAAGGCAAAATACAGATGAACATGACGAGTAAATGCTTGGTGGGATTATGGCTAT 300
 QY 610 GlnIleAlaGlyMetAlaAsnArgThrLeuValAspAsnAsnThrTrpAsnAsnSer 629
 Db 301 CAGATAGCTGGAATGGCTATAGAACATACCTGGTGAATAAATACACCTGGATACAGC 360
 QY 630 HisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMet 649
 Db 361 CACATAGCACTGGTGGGAAGCTATGCTTCTTAATGAACACACAGCCTATAAATCATG 420
 QY 650 ArgThrLeuAspValAspTyrValLeuValIlePheGlyClyValIleGlyTrpSerGly 669
 Db 421 AGGACTCTAGATGATAGATATGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 480
 QY 670 AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp 689
 Db 481 GATGATATCAACAAATTTCTCTGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 540
 QY 690 IleArgGluSerAspTyrPheThrProGlnGlyLysPheArgValAspLysAlaGlySer 709
 Db 541 ATTCGGGAAGTGCATATTTTACCCACAGGAGGAATTCCTGTAGACACAGGAGGAGGAGG 600
 QY 710 ProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGln 729
 Db 601 CCTACTTTGTTGAAATTTGCTTATGATAAATGTCATCTACATGAGATTTGGAGAAATGCAG 660
 QY 730 LeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLys 749
 Db 661 CTGGATTTTCTGACACCCCAAGGTTTGGCCGACAGGATTCGCTAGACAAAGCAGGATCC 720
 QY 750 AspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIle 769
 Db 721 GACATTAATTCMAACATTTGGAAGAGCCTTTACATCAGACACTGGCTTGTAGGATA 780
 QY 770 TyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAsn 789
 Db 781 TATAAAGTAAGAACGACCTGTAAACAGGGAGACATTAGATCACAACCTCGAGTCACCAAC 840
 QY 790 IlePheProLysGlnLysTyrLeuSerLysLysThrLysArgLysGlyTyrIle 809
 Db 841 ATTTTCCCAACACAGAGTATTTGTCAAAGAGACTACCAAAAGGAAGGCTGGCTACATT 900
 QY 810 LysAsnLysLeuValPheLysGlyLysLysLysLysSerLysLysThrVal 826
 Db 901 AAAAATAAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAGAGAGCTGTT 951

RESULT 11
 BD123520 2510 bp DNA linear PAT 18-SEP-2002
 LOCUS Secretory protein or membrane protein.
 DEFINITION
 ACCESSION BD123520
 VERSION BD123520.1 GI:23218465
 KEYWORDS JP 2002017376-A/29.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Ota.T.; Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
 Hayashi,K.
 SECRETORY PROTEIN OR MEMBRANE PROTEIN
 TITLE Secretory protein or membrane protein
 JOURNAL Patent: JP 2002017376-A 29 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 COMMENT OS Homo sapiens (human)
 PN JP 2002017376-A/29
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253173
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, TOMOYASU
 PI SUGIYAMA,
 PI KOJI HAYASHI
 PC

C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
 10, C12P21/02, C12Q1/68//C12P21/08, C12N15/00, C12N5/00 CC
 Secretory protein or membrane protein
 FH Key Location/Qualifiers
 FT CDS (94)...(951).
 FEATURES
 source 1..2510
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores: 0 Length: 2510
 Pred. No.: 317.00 Matches: 317
 Score: 317.00
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.38% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-2 (1-826) x BD123520 (1-2510)
 QY 510 LeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLysThrGluGlu 529
 Db 1 TTGTATGATAAGCAGGTAAGTGAAGAAACATGCAACTGAACAGGAAAAAAGTCAAGAG 60
 QY 530 GlyLeuGlyProAsnLysSerIleValThrMetLeuMetLeuMetLeuMetMet 549
 Db 61 GGATTAGGCCCTAATATAAAGCAATGTCACCATGTTGATGCTGATGCTATGATGATG 120
 QY 550 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeu 569
 Db 121 TTGTGTCCTCCTGTCACCTGGGTGCAACCAATGCTACTCTAGTCCAAGTGTAGTCTG 180
 QY 570 AlaserTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPhe 589
 Db 181 GCCTCATCATCATGATGCGACCAAGGATATCTTAGATGATTTAGAGAGCTTACTTT 240
 QY 590 TrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyr 609
 Db 241 TGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTCTGTGGGATTATGGCTAT 300
 QY 610 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 629
 Db 301 CAGATAGCTGGAATGGCTAATAGAACTAGTGGTGGATATAACACCTGGAAATAACAGC 360
 QY 630 HisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMet 649
 Db 361 CACATAGCACTGGTGGGAAGCTATGCTTCTTAATGAACACAGGAGGAGGAGGAGGAGG 420
 QY 650 ArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 669
 Db 421 AGGACTCTAGATGATAGATATGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 480
 QY 670 AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp 689
 Db 481 GATGATATCAACAAATTTCTCTGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 540
 QY 690 IleArgGluSerAspTyrPheThrProGlnGlyLysPheArgValAspLysAlaGlySer 709
 Db 541 ATTCGGGAAGTGCATATTTTACCCACAGGAGGAATTCCTGTAGACACAGGAGGAGGAGG 600
 QY 710 ProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGln 729
 Db 601 CCTACTTTGTTGAAATTTGCTTATGATAAATGTCATCTACATGAGATTTGGAGAAATGCAG 660
 QY 730 LeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLys 749
 Db 661 CTGGATTTTCTGACACCCCAAGGTTTGGCCGACAGGATTCGCTAGACAAAGCAGGATCC 720
 QY 750 AspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIle 769
 Db 721 GACATTAATTCMAACATTTGGAAGAGCCTTTACATCAGACACTGGCTTGTAGGATA 780
 QY 770 TyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAsn 789
 Db 781 TATAAAGTAAGAACGACCTGTAAACAGGGAGACATTAGATCACAACCTCGAGTCACCAAC 840
 QY 790 IlePheProLysGlnLysTyrLeuSerLysLysThrLysArgLysGlyTyrIle 809
 Db 841 ATTTTCCCAACACAGAGTATTTGTCAAAGAGACTACCAAAAGGAAGGCTGGCTACATT 900
 QY 810 LysAsnLysLeuValPheLysGlyLysLysLysLysSerLysLysThrVal 826
 Db 901 AAAAATAAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAGAGAGCTGTT 951

```

FEATURES             Location/Qualifiers
     source                1..2546
                        /organism="unidentified"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:32644"
ORIGIN
Alignment Scores:
Pred. No.:              0           Length:      2546
Score:                 323.00      Matches:      323
Percent Similarity:    100.00%     Conservative: 0
Best Local Similarity: 100.00%     Mismatches:  0
Query Match:          39.10%      Indels:      0
DB:                   6           Gaps:         0

US-10-028-384-2 (1-826) x BD063986 (1-2546)
QY 504 LysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGlu 523
Db 1 AAAGAAACCAGGAAATTTGTATGAAGCAGGTAAAGTGAGGAAACATGCAACTGAA 60
QY 524 GlnGluThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMet 543
Db 61 CAGGAAAAAATCGAAGAGGATAGGCCCTAATAATAAAAGCATTTGTCACCATGTTGATG 120
QY 544 LeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSer 563
Db 121 CTGATGCTATTGATGATGTTTGGTGTCCACTGACCTGGGTCAAGCAATGCTACTCT 180
QY 564 SerProSerValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAsp 583
Db 181 AGTCCAGTGTAGTCTCGGCTCATACATCATGCGCACCGAGGAATATCTTAGATGAT 240
QY 584 PheArgGluAlaTyrPheTrpLeuArgAlaAsnThrAspGluHisAlaArgValMetSer 603
Db 241 TTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACGAGTAATGCT 300
QY 604 TrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsn 623
Db 301 TGGTGGGATATTGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGGTGGATAAT 360
QY 624 AsnThrTrpAsnSerHisAlaLeuValGlyLysAlaMetSerSerAsnGluThr 643
Db 361 AACACCTGGATTAACAGCCCATAGCACTGGTGGAAAGCATATGCTTCTTAATGAACA 420
QY 644 AlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGly 663
Db 421 GCAGCCTATAAATCATGAGCACTCTAGATGATGATTATGTTTGGTTATTTTGGAGGG 480
QY 664 ValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGlu 683
Db 481 GTTATGGCTATTCTGGGATGATATCAACAAATTTCTCGATGGTGTAGGATAGCTGAA 540
QY 684 GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
Db 541 GGAGAAATCCCAAGACATTCGGGAAAGTGACTATTTTACCCACAGGAGAAATTCCT 600
QY 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
Db 601 GTAGCAAAAGCAGGATCCCTTCTTGTGAATTCCTTATGATTAATGATCACTACTAC 660
QY 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn 743
Db 661 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCACCGAACACGTAAT 720
QY 744 AlaGluIleGlyAsnLysAspIleValPheIleHisLeuGluGluAlaPheThrSerGlu 763
Db 721 GCTGAGATTGAATAGGACATTAATTAACACATTTGGAAGACCTTTACATCAGAA 780
QY 764 HisTrpLeuAlaGlyTyrLysValIleAlaProAspAsnArgGluThrLeuAspHis 783
Db 781 CACTGGCTTGTAGGATATAAAGTAAAGACCTGATAACAGGGAGACATTAGATAC 840

```

```

QY 784 LysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrLys 803
Db 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATATTTCCTCAAGAGACTACCAA 900
QY 804 ArgLysArgGlyTyrIleLysAsnLysLeuValPheLysGlyLysLysLysSerLys 823
Db 901 AGGAAGCGTGGCTACATTAAATAAATAGCTGGTTTTTAAAGAAAGCGAAGAAATATCTAAG 960
QY 824 LysThrVal 826
Db 961 AAGACTGTT 969
RESULT 10
AX136135
LOCUS AX136135
DEFINITION Sequence 57 from Patent EP1067182.
ACCESSION AX136135
VERSION AX136135.1 GI:14272543
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Oca, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and
Hayashi, K.
TITLE Secretory protein or membrane protein
JOURNAL Patent: EP 1067182-A 57 10-JAN-2001;
Helix Research Institute (JP)
FEATURES
source 1..2510
location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
94..954
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC39710.1"
/db_xref="GI:14272544"
/translation="MLMLMLMFAVHCTTWTNSAVSSPVLSVSYNHDGTRNLLDDP
REAFVRLONTDEHARVMSWDYGYIAGMANRTTLDVNTNNSHIALYVKGAMSSNE
TAAVKLMRLDLDVYLVIFGGVIGSGDDINKFLMWRVIAEGEPKDIRESDYTPQG
EFKVDKAGSFTLLCLMYMSYRFGEMQLDFRTPGPDRTNNABIGNDKIKFKHLEE
AFTSEHLVRIYKVKAPDNRLDHPKVTWTFIPKQYLSKKTTRKRGYIKNKLVEK
KGKKISKKTV"
ORIGIN
Alignment Scores:
Pred. No.:              0           Length:      2510
Score:                 317.00      Matches:      317
Percent Similarity:    100.00%     Conservative: 0
Best Local Similarity: 100.00%     Mismatches:  0
Query Match:          38.38%      Indels:      0
DB:                   6           Gaps:         0

US-10-028-384-2 (1-826) x AX136135 (1-2510)
QY 510 LeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnLysThrGluGlu 529
Db 1 TTGTATGATAGGACAGGTAAAGTGAGAAACATGCACTGAACAGGAAAAAACTGAAGAG 60
QY 530 GlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMet 549
Db 61 GGATTAGGCCCTAATAATAAAAGCATTTGTCACCATGTTGATGCTGATCTATTGATGATG 120
QY 550 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValVal 569
Db 121 TTTGCTGTCCACTGTACCTGGGTGTCACAGCAATGCTACTCTAGTCCAAAGTGTAGTCTGT 180
QY 570 AlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPhe 589
Db 181 GCCTCATACATCATCATGATGCGACCGAGGAATACTTTAGATGATTTTATAGAGAGCTTACTTT 240

```

QY 712 uLeuAsnCysLeuMetTyrIysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspH 732
 Db 2203 GTTAACTGCCCTTATGTATATAATGTCTACTACAGATTGGAGAAATCGACTAGATT 2262
 QY 732 eArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleY 752
 Db 2263 TCGCACTCCCCAGGCTTTTGCCGACACAGTATATCTGAGATTGGAAATTAAGACATTA 2322
 QY 752 sPheLysHisLeuGluLulAlaPheThrSerGluHisTrpLeuValArgIleTyrIysVa 772
 Db 2323 ATTCAAGCATTTGGAGGAAGCTTTTACATCAGAGCACTGGCTTGTGAGGATATATAAGT 2382
 QY 772 llysalapProAspAsnArgGluThrLeu 781
 Db 2383 GAAAGCACCTGACACACAGGGAGACACTA 2410
 RESULT 8
 AX099510
 LOCUS
 DEFINITION Sequence 150 from Patent WO0119988.
 ACCESSION AX099510
 VERSION AX099510.1 GI:13538588
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 Jacobs, K., Mccoy, J.M., Lavallie, E.R., Collins-Racie, L.A., Evans, C.,
 Merberg, D., Treacy, M., Bowman, M.R., Spaulding, V., and Agostino, M.J.
 Secreted proteins and polynucleotides encoding them
 Patent: WO 0119988-A 150 22-MAR-2001;
 Genetics Institute, Inc. (US)
 FEATURES
 source
 1. 2546
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 2546
 Score: 323.00 Matches: 323
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 29.10% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-2 (1-826) x AX099510 (1-2546)
 QY 504 lYsArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGlu 523
 Db 1 AAAAGAAACCAAGGAATTTCTATGATAGAGCAGGTAAAGTGAGGAAACATGCAACTGAA 60
 QY 524 GlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMet 543
 Db 61 CAGGAAAAAATCGAAGAGGGATTAGGCCCTTAATAAAAGCAATGTCCACCATGTGATG 120
 QY 544 LeuMetLeuLeuMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSer 563
 Db 121 CTGATGCTATTGATGATGTTTGTCTCCACTGACCTGGGTCCACAGCAATGCTACTCT 180
 QY 564 SerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAsp 583
 Db 181 AGTCAAGGTGAGTCTGGCTCTATCAATCATGATGCGACCGAGGAATCTTAGATGAT 240
 QY 584 PheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSer 603
 Db 241 TTTAGAGAGCTTACTTTTGGCTTAAGGCAAAATACAGATCAACATGCAAGTAAATGCTCT 300
 QY 604 TrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsn 623

Db 301 TGGTGGGATTATGGCTATCAGATAGCTGGATGGCTAATAAGAACTACGTTGGTGGTAAT 360
 QY 624 AsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThr 643
 Db 361 AACACCTGGAATAACAGCCACATAGCACTGGTGGGAAAAGCTATGCTCTTCTTAATGAACA 420
 QY 644 AlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGly 663
 Db 421 GCAGCCTTAATAATCATGAGGACTCTAGATGATGATTAATGTTTGGTTATTTTGGAGG 480
 QY 664 ValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGlu 683
 Db 481 GTTATTGGCTATTCTGGTGATGATCAACAAATTTCTCTGGATGGTGTAGGATAGCTGAA 540
 QY 684 GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
 Db 541 GGAGAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCACAGGGAGAAATTCGGT 600
 QY 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
 Db 601 GTAGACAAAGCAGGATCCCTACTTTTGTGTAATTCCTTATGCTATAAAATGTCTACTAC 660
 QY 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn 743
 Db 661 AGATTTGGAGAAATGCGCTGGATTTTGTACACCCCGGGTTTGTACCGACACGTAAT 720
 QY 744 AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGlu 763
 Db 721 GCTGAGATTGGAATTAAGGACATTAATAATTCAAACATTTTGGAAAGAGCCCTTTACATCA 780
 QY 764 HisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThrLeuAspHis 783
 Db 781 CACTGCTTGTAGGATATATTAAGTAAGCACTGTATACGGGAGACATTAAGATCAC 840
 QY 784 LysProArgValThrAsnIlePheProLysGlnLysTyrLysLeuSerLysLysThrLys 803
 Db 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGTATTTGTCAAGAAGACTACCAAA 900
 QY 804 ArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLysIleSerLys 823
 Db 901 AGGAAGCGTGGCTACATTAATAATAAGCTGGTTTTTAAGAAGGACAGAAATATCTAAG 960
 QY 824 LysThrVal 826
 Db 961 AAGACTGTT 969

RESULT 9

BD063986

LOCUS

DEFINITION

SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM.

ACCESSION

BD063986

VERSION

BD063986.1

KEYWORDS

JP 2001506848-A/10.

SOURCE

unidentified

ORGANISM

unclassified.

REFERENCE

1 (bases 1 to 2546)

AUTHORS

JACOBS, K., MCCOY, J.M., LAVALLIE, E.R., RACIE, L.A., MERBERG, D.,

TREACY, M., SPAULDING, V. AND AGOSTINO, M.J.

TITLE

SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

JOURNAL

PATENT: JP 2001506848-A 10 29-MAY-2001;

COMMENT

GENETICS INSTITUTE INC

PN JP 2001506848-A/10

PD 29-MAY-2001

PF 12-DEC-1997 JP 1998525996

PR 13-DEC-1996 US 08/766263

PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI

DAVID MERBERG,

PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC

C12N15/12, C12N5/10, C07K14/47, A61K38/17

CC Strandedness: Double;

CC Topology: Linear;

FH Key

Location/Qualifiers.

1069	DB	CGCTTTTTCGATGATCTCGAGACCGGTTGCAAAAACAGAGGTCAGACCCCTTTCTT	1128
354	QY	eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrTh	374
1129	DB	TTTGGGTGTCCTACTAGCTGCAGGGCGTGTGTTCCCTTAGTGTCACTATCTGCATACAC	1188
374	QY	rGlyTyrIleAlaProTyrSerGlyArgPheTyrSerLeuTyrAspThrGlyTyrAlaLys	394
1189	DB	AGGTATATTGTCACCATGAGTGGCAGGTTTATTCACTATGGGATACTGGGTATGCAAA	1248
394	QY	sIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrValSerPh	414
1249	DB	AATACACATTCCAATTATTGCATCAGTGTCTGAACATCAGCCTACGACATGGGTGTCTTT	1308
414	QY	ePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTyrPheCysIleLys	434
1309	DB	CTTTCTTTGATCTACATATTCTTGTATGTACCTTCCAGCAGCGCTATGGTCTGTCAATCAA	1368
434	QY	sAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAl	454
1369	DB	AAATATCAATCAATGAAAGAGTATTGTGGCTGTGTATCGCATCATCTGTGTACTTTGC	1428
454	QY	aGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaI	474
1429	DB	CGAGTGTATGGTGGCGCTGATGCTGACTCTGAACCCCGGTCTGTGCATGTGTGCGCCAT	1488
474	QY	eAlaPheSerAsnValPheGluHisTyrLeuGlyAspAspMetLysArgGluAsnProPr	494
1489	DB	CGCTCTCTCCAATGTGTTTTGAGCAGTATTGTGGGGATGACATGAAAAAGGAAACCCACC	1548
494	QY	oValGluAspSerSerAspGluAspAspLysArgGAsnGln-GlyAsnLeuTyrAspLysA	514
1549	DB	TGTGGAGGACGAGCGTGTATGAGGATGCANAAAGAAACCC-AGGAAATCTTGTATGACNAGG	1607
514	QY	LaGlyLysValArgLysHisAla-ThrGluGlnGluLysThr-GluGluGlyLeuGlyPr	533
1608	DB	CAGGTAAAGTGAGGAAGCATGTG-ACAGAGCAAGAGAA-ACCTCAAGAGGGCTTGGGCC	1665
533	QY	oAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHi	553
1666	DB	CAACATCAAAAGCATTTGTGACCATGCTGATGCTCATGCTCTGATGATGTTCGGGTCCA	1725
553	QY	sCysThrTyrValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAs	573
1726	DB	CTGCAGCTGGGTCAACAAGCAAGCCCTACTCCAGTCCAAGTGTGCTCTGCCCTCTACAA	1785
573	QY	nHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTyrLeuArgGl	593
1786	DB	TCATGATGTTACCAAGGATATATTAGATGATTTTAGAGAAGCGTACTTTTGGCTGAGACA	1845
593	QY	nAsnThrAspGluHisAlaArgValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGl	613
1846	DB	AAACACCGGATGAACACGCCCGGCTCATGTCTGTGTGGGACTACGGCTATCAGATTGCTGG	1905
613	QY	yMetAlaAsnArgThrThrLeuValAspAsnAsnThrTyrAsnAsnSerHisIleAlaLe	633
1906	DB	CATGGCCAAACAGACCCTCTGGTGGATAACAACACCTGGAACAACAGCCACATCGCACT	1965
633	QY	uValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr-LeuA	653
1966	DB	GGTCGGAAAAGCATGTGCTTCCAAATGAACCGCCCTATTAAAAATCATGAGGTC-CTTGT	2024
653	QY	spValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleA	673
2025	DB	ATGTCGATTTATGTGTTGGTTATTTTCGGAGGAGTATTGGCTATTCCGGGACGATATCA	2084
673	QY	snLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluS	693
2085	DB	ACAAGTTCTCTGGATGGTCAGATAGTGTGAAGGGGAGCATCCCAAGAATCCCGGAAG	2144
693	QY	er-AspTyrPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThrLe	712
2145	DB	G-TGACTATTTTAC-CCACAGGAGAGTTCCGACTAGACAAAGCTGGGTCTCTACTCT	2202

FEATURES
source

Location/Qualifiers
1..1664
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEMBA1006430"
/tissue="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/note="cloning vector: pME189FL3"
28..1362
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC11076.1"
/db_xref="GI:22760123"
/translations="MAEPSPAPESKHSKLSNPSWGLMALGNRHHGHPGQAQAHKA
AGGAAPKPPAPAGLSGSLQSPAGWQLLFTILFLAWLAGFSRLFAVIRFESIHIF
DPWFNYSRTHLASHGFEYELNWFDERAYPLGRIVGGTYFPLMITAGLHILNLT
NITVIRDCVCFVLAFTPSGLTSISTELLRELWNOGALLAACFIAIVPGVTSRVAG
SFDNEGIAI PALOPTYLVKVSFKTGVFWCCCLSVFYMVSANGGVFVFNILPLH
AFVLLLMQRYSKRVLYAYSTFYIVGLILSMQIPFVGFPQIRTSHEMAAGVFAILLQAY
AFLOYLDRDTKQEFOTLFLGYSLAAGAVELSVLYLVTVGYIAPWGRFVSLWDVGY
AKIHPIIASVSEHQFTWVSFPFDLHILVCFPAGLMFCIKXNDERVFGA"

CDS

568 ACATCTATATCTACTTTCTCTGCTTCAAGAGAACTTTTGGAAACCAAGAGAGAGACTTTT 627
201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
628 GCTGCTTTTATTGCTATTGTACCAGGCTACATATCTCGGTGAGTAGCTGGATCTTT 687
221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValIleSer 240
688 GATAAAGAGGAGGATGCTATTTTTCAGCTTTCAGTTCATCATCTATTTATGGGTAATCT 747
241 ValIleThrGlySerValPheThrMetCysCysLeuSerTyrPheTyrMetVal 260
748 GTAAAAAAGTGGGTGAGTCTTTTGGCAATGTGCTGCTTATCTATTTCTATATGTC 807
261 SerAlaTyrGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
808 TCTGCTTGGGTGGTATTGATTTATCATCAATCTTATTCACCTGCATGATTTTGTGTG 867
281 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
868 TTACTCATGCAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTGTG 927
301 GlyLeuIleIleSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
928 GGTTTAATATTATCAATGCAGATACCTTTTGGGATTTCCAGCAATCAGAACAAAGTAA 987
321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
988 CACATGGCAGCTGCAGGTGCTTTTCATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1047
341 ArgAspArgLeuThrIleGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
1048 AGAGACCGATTAAACAAACAGAGTTCAGACCCCTTTCTTTTGGGTGATCACTAGCT 1107
361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyr 380
1108 GCAGGTGCTGTGCTTCTTAGTGTCTATTTGATCTTATACAGGTTACATTGCACCATGG 1167
381 SerGlyArgPheTyrSerLeuThrPheThrGlyTyrAlaIleHisIleProIleIle 400
1168 AGTGGCAGGTTTATTCTATTGTTGGGATCTGGGTATGCAAAATACATTTCCAAATATT 1227
401 AlaSerValSerGluHisGlnProThrThrTyrValSerPhePheAspLeuHisIle 420
1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTCTTTCTTCTTGTATCATATT 1287
421 LeuValCysThrPheProAlaGlyLeuTyrPheCysIleIleLysAsnIleAsnAspGluArg 440
1288 CTGTGTATGATACCTTCCAGCAGGCTTTGGTCTGCTATCAAAATATCAAGATGAAGA 1347
441 ValPhe 442
1348 GTATTT 1353

US-10-028-384-2 (1-826) x AK074587 (1-1664)

ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 1664
Score: 442.00 Matches: 442
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.51% Indels: 0
DB: 9 Gaps: 0

Qy 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTyr 20
Db 28 ATGGCGGAGCCCTCGGCGCGGAGAGCAAGCAAGTCTCTCAACTCGTCCCGGTGG 87
Qy 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
Db 88 AGTGGCCTCATGGCCCTGGGAACAGCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 147
Qy 41 AlaHisLysAlaAlaGlyGlyAlaAlaProGlyProAlaProAlaProAlaProGly 60
Db 148 GCGCACAGGCGGCG 207
Qy 61 GlyLeuSerGlnProAlaGlyTyrGlnSerLeuSerPheThrIleLeuPheLeuAla 80
Db 208 GGGCTGTGCG 267
Qy 81 ThrLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100
Db 268 TGGCTTGGCGGCTTCAGTCCGCGCTTTCGCGCTCATCGCTTCGAAAGCATCATCCAC 327
Qy 101 GluPheAspProTyrPheAsnTyrArgSerThrHisLeuAlaSerHisGlyPheTyr 120
Db 328 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTTGCATCTCATCGGTTCAT 387
Qy 121 GluPheLeuAsnTyrPheAspGluArgAlaTyrTyrProLeuGlyArgIleValGlyGly 140
Db 388 GAATTTTAAATGTTTGAATGAAGAGCATGATGATCCACTAGGAAGATAGTAGTGGT 447
Qy 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTyrIleLeuAsnThrLeu 160
Db 448 ACTGTTTACCAGGTTGATGATAACCGCTGGCGCTTATTCAATGGATTTTAAATACATG 507
Qy 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
Db 508 AACATACTGTTTACATAAGAGAGCGTGTGTGTCTTCTGACCACTTTTAGCGGCGCT 567
Qy 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTyrAsnGlnGlyAlaGlyLeuLeu 200

RESULT 7

AX799084 2710 bp mRNA linear PAT 08-OCT-2003
LOCUS Sequence 3 from Patent WO03054008.
DEFINITION AX799084
ACCESSION AX799084
VERSION AX799084.1 GI:37605059
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Perreault, C. and McBride, K.
TITLE Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy
JOURNAL Patent: WO 03054008-A 3 03-JUL-2003;
FEATURES Competigene Inc. (CA)
Location/Qualifiers
1..2710

Primer for synthesizing full-length cDNA and use thereof FH Key

Location/Qualifiers
(28)...(1359).Location/Qualifiers
1..1664/organism="Homo sapiens"
/mol_type="genomic DNA"

/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 1664

Score: 442.00 Matches: 442

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 53.51% Indels: 0

DB: 6 Gaps: 0

US-10-028-384-2 (1-826) x BD127193 (1-1664)

QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAenSerSerProTTP 20

Db 28 ATGGCGGAGCCCTCGGCCCGGAGAGCAAGTCCTCAACTCGTCCCGTGG 87

QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40

Db 88 AGTGGCCTCATGGCCCTGGGAAACAGCGGACCGCCACCGCGCGCGCGCGCGCG 147

QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProlLysProAlaProAlaGlyLeuSerGly 60

Db 148 GCGCACAGGCG 207

QY 61 GlyLeuSerGlnProAlaGlyTTPGlnSerLeuLeuSerPheThrLeuPheLeuAla 80

Db 208 GGGCTGTGCGACCG 267

QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100

Db 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTTCATCGCTTCGGAAGCATCCAC 327

QY 101 GluPheAspProTTPPheAsnTyArgSerThrHisHisLeuAlaSerHisGlyPheTy 120

Db 328 GAGTTCGACCGCGTGTAACTATAGATCAACATCATCTTCGATCTCATGGTGTCTAT 387

QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTyProLeuGlyArgIleValGlyGly 140

Db 388 GAATTTTAAATTGTTGATGAAGACATGGTATCCACTAGGAAGATAGTAGTGT 447

QY 141 ThrValTyProGlyLeuWeIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160

Db 448 ACTGTTTACCGAGGTTGATATACCGCTGCTTATTCATTCGATTTTAAATACATG 507

QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180

Db 508 AACATAACTGTTTACATAAGAGACGTGTGTGTCTTCCTGACCAACTTTTAGCGCCT 567

QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200

Db 568 ACATCTATCTACTCTTCCTGCTTACAGAGAACTTTTGGACCAAGGAGGAGGACATTTA 627

QY 201 AlaAlaCysPheIleAlaIleValProGlyTyIleSerArgSerValAlaGlySerPhe 220

Db 628 GCTGCTGTTGTTTATTTGCTATGTACCGCTTACATATCTCGGTTCAGTAGCTGCTTT 687

QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyTrpLeuTrpValIleSer 240

Db 688 GATATGAGGAGGATTTGCTATTTTTCACCTTCAGTTTCATACACTATTTATGGTAAATCT 747

QY 241 ValIleThrGlySerValPheTrpThrMetCysCysLeuSerTyTrpPheTyMetVal 260

Db 748 GTAAACATGGCTGCTGCTTTTGGACATGTGCTGCTTATCTCTATTTCTATATGCTC 807

QY 261 SerAlaTrpGlyGlyTyTrpValPheIleIleAsnLeuIlePheLeuValPheValLeu 280

Db 808 TCTGCTTGGGGTGGTGTATGTATTATCAATCAATCTTATTCCACTGCATGTATTGTGTG 867

QY 281 LeuLeuMetGlnArgTySerLysArgValTyIleAlaTySerThrPheTyIleVal 300

Db 868 TTACTGATGCAGAGATACAGCAAGAGTCTACATAGCATATAGCACCTTCTTACATTGTG 927

QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320

Db 928 GGTTTAATATATCAATGTCAGATACCTTTTGGGATCCAGCCCAATCAGAACAGTGA 987

QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyAlaPheLeuGlnTyIle 340

Db 988 CACATGCGAGCTGCGAGTGTCTTTGCAATGCTGCAAGCTTATGCTTCTTTCAGTATCTG 1047

QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360

Db 1048 AGAGACCGATTAAACAAACAGAGTTCAGAGCCCTTTCTTTTGGGTGTATCATAGCT 1107

QY 361 AlaGlyAlaValPheLeuSerValIleTyLeuThrTyTrpGlyTyIleAlaProTTP 380

Db 1108 GCAGTGTCTGTGTTCTTCTAGTGTCACTATTGACATTATACAGGTTACATTGCCACATG 1167

QY 381 SerGlyArgPheTySerLeuTrpAspThrGlyTyAlaIleIleHisIleProIleIle 400

Db 1168 AGTGGCAGGTTTATTCATTTGGGATCTGGGTATGCAAAATACACATTCCAATTATT 1227

QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIle 420

Db 1228 GCATCAGTGTCTGAGCATCAACCTACGACTGGGTGCTCTTCTTCTTGTGATCATATT 1287

QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleIleAsnIleAsnAspGluArg 440

Db 1288 CTGTATGTACCTTCCAGAGGCTTTGGTCTGCATCAAAATATCAACGATGAAGA 1347

QY 441 ValPhe 442

Db 1348 GTATTT 1353

RESULT 6

AK074587

LOCUS

DEFINITION Homo sapiens cDNA FLJ90106 fis, clone HEMBA1006430, weakly similar to Human putative transmembrane protein precursor (B5) mRNA.

ACCESSION AK074587

VERSION AK074587.1 GI:22760122

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

2 Isogai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T., Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S., Kawai-Hio, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Kojima, S., Nagahari, K., Masuho, Y., Ota, T., Okano, K., Yoshikawa, Y., Aotsuka, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and Ninomiya, K.

3 NEDO human cDNA sequencing project

4 Unpublished

5 Isogai, T. and Otsuki, T.

6 Direct Submission

7 Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

8 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).

```

/codon_start=1
/protein_id="CAP85831.1"
/db_xref="GI:45502339"
/translation="MSEPSAPESKHSKSLNSPWSGLMAGNSRHHGHHGAGCAHKA
AGAAPKPAAGSGLSQFQWQSLLSFTILFLAWLAGFSRFLPAVIRFESIIEF
DFWNYRSTHLAGHGYEFLNMFDERAMYPGLRIVGGTVYFGLMITAGLIHWIINTL
NFDNIGAI PALQFTYLYLWVKVSKTGVFVTCCLSYFVMSAWGVYFIINLPLH
FVLLLMCRYSKRVYIAYSTFYLVGLISMCIPVGFQPIRIRSEHMAAGVALLQAY
AFQYLRDLRTKQRFQTLFPLGSLAGAVFLSVILTYTGHIAPWSRFPYSLNDY
AKIHPIIASVSEHQPTTWSFFDLHLVCTIPAGLWFCIKMINDEVRFGA"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 0 Length: 1664
Score: 442.00 Matches: 442
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.51% Indels: 0
DB: 6 Gaps: 0

```

US-10-028-384-2 (1-826) x CQ782554 (1-1664)

```

QY 1 MetAlaGluProSerAlaProGluSerLysHisGlySerSerLeuAsnSerSerProTTP 20
DB 28 ATGGGGGAGCCCTCGGCCCGCGAGAGCAAGATCGTCCCTCAACTCGTCCCCGTGG 87
QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
DB 88 AGTGGCCCTCATGGCCCTGGGAAACAGCGCGCACGCCACCCAGCGCGGCCCGCCAGTGC 147
QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
DB 148 GGCACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 207
QY 61 GlyLeuSerGlnProAlaGlyTyrPglNserLeuLeuSerPheThrIleLeuPheLeuAla 80
DB 208 GGGCTGTCCGACCGCGCTGGGTGGCGTGGCTTCTCTCCCTTCCCTTCTCTCTCTCT 267
QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis 100
DB 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCCCGTCTATCCGCTTCGAGAGCATCATCCAC 327
QY 101 GluPheAspProTTPPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyr 120
DB 328 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTTGCATCTCATGGGTCTAT 387
QY 121 GluPheLeuAsnTTPPheAspGluArgAlaTyrProLeuGlyArgIleValGlyGly 140
DB 398 GAATTTTAAATGGTTTGTATGATCAAGAGCATGTATCCACTAGAGAGATAGTAGTGT 447
QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
DB 448 ACTGTTTACCCAGGCTTCATGATACCGCTGGCCCTTATTCATGGATTTTAAATACATG 507
QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
DB 508 AACATACTGTTCATCAAGAGAGCGTGTGTGTCTTCCGACCACTTTTACGGCCCTT 567
QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
DB 568 ACATCTATATCTTCTTCTCTTACAGAGAGACTTTTGGAAACCAAGAGAGAGACTTTTA 627
QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
DB 628 GCTGCTGTGTATGTATGTATGTACAGGCTACATATCTCGGTTCAGTGTATCTTCTT 687
QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSer 240
DB 698 GATAATGAGGATTCGCTATTTTTCGCTTCAGTTCATATCTTATGGGTAAATCT 747
QY 241 ValLysThrGlySerValPheTyrThrMetCysCysLeuSerTyrPheTyrMetVal 260

```

```

DB 748 GTAAABAACGGGTGTCAGTCTTTTGGCAAAAGTGTCTGCTGCTATCTATCTATATGTC 807
QY 261 SerAlaTyrGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
DB 808 TCTGCTTGGGTGGTGTATGTATTTATCATCAATCTTATTCACATGCTATGTTGTGTG 867
QY 281 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
DB 868 TTACTGATGCAGAGATACAGCAAGAGCTACATAGCATATAGCATCTTCTACATTTG 927
QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
DB 928 GGTTTAATATTAATCAATGCAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTAA 987
QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
DB 988 CACATGGCAGCTGCAGGTGCTTTTCATTTGCTTGCAGCTTATGCTTTCTTCAGTATCTG 1047
QY 341 ArgAspArgLeuThrLysGlnGlnPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
DB 1048 AGAGACCGATTAACAAACCAAGAGTTCAGAGCCCTTTCTTTTGGGTGTATCACTAGCT 1107
QY 361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTTP 380
DB 1108 GCAGGTGCTGTCTTCTTGTATGTCATCTATTGCTTATACAGGTTCATATGCACATGG 1167
QY 381 SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIle 400
DB 1168 AGTGGCAGGTTTTTATTCATTTGGGATATCGGTATGCAAAATAACATCCATTAATT 1227
QY 401 AlaSerValSerGluHisGlnProThrThrTyrValSerPhePheAspLeuHisIle 420
DB 1228 GCATCAGTGTCTGAGGATCAACTACGATTCGGGTGCTTCTTCTTGTATCTACATATT 1287
QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
DB 1288 CTTGTATGATACCTTCCAGCAGCGCTTTGGTCTCTGTCATCAAAATAATCAACGATGAA 1347
QY 441 ValPhe 442
DB 1348 GTATTT 1353

```

RESULT 5
BD127193
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BD127193 1664 bp DNA linear PAT 18-SEP-2002
Primer for synthesizing full-length cDNA and use thereof.
BD127193
BD127193.1 GI:23222138
JP 2002017375-A/2624
Homo sapiens (human)

REFERENCE
AUTHORS
Oka, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
Koga, H.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A/2624
HBLIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/2624
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OKA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII
PI YUKI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINTCHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
'10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC

QY 195 GlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArg 214
 DB 939 CAGAGAGCAGGACCTCTAGCTGCTTCTTCTATCTGATCTGACAGGAGTACATATCTCGG 998
 QY 215 SerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 234
 DB 999 TCAGTGGGGGATCCTTTGATAATGAGGCATTCGCCATTTTGGCGCTTCACTTAC 1058
 QY 235 TyrLeuTyrValIleSerValIleThrGlySerValPheThrMetCysCysLeu 254
 DB 1059 TACTTATGGGTAAAGTCTGTGAGACCGGCTCTGTCTTGGACAAAGTCTGCTGCTTG 1118
 QY 255 SerTyrPheTyrMetValSerAlaTyrPheGlyTyrValPheIleIleAsnLeuIlePro 274
 DB 1119 TCATATTTCTACATGCTCTCTCGGTGGGAGGTTATGTGTTCATCATCAACCTCATCCCT 1178
 QY 275 LeuHisValPheValLeuLeuLeuMetGlnArgTyrSerIleValTyrIleAlaTyr 294
 DB 1179 CTCATGTGTGTGCTGCTCTGATGCAGAGGTACAGAGAGTCTACATAGCATAT 1238
 QY 295 SerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGln 314
 DB 1239 ACCACTTTCTCATTTGGGTTTAAATATTCATTCAGATACCTTTTGTGGGATTCAG 1298
 QY 315 ProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyr 334
 DB 1299 CCAATCAGACACAGCGAGCAGCATGGCAGCTGCAGGTGCTTTGGCTGCTGCAAGCTTAC 1358
 QY 335 AlaPheLeuGlnTyrLeuArgAspArgLeuThrIleGlnGluPheGlnThrLeuPhePhe 354
 DB 1359 GCTTTTTCAGTATCTCAGAGACCGGTTGACAAACAGGAGTTCACAGACCTTTTCTTT 1418
 QY 355 LeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThr 374
 DB 1419 TTGGGTGCTCCTACCTAGCTGCAGGCGCTGTGTCTTCTAGTCTATCTGACATACACA 1478
 QY 375 GlyTyrIleAlaProThrSerGlyArgPheTyrSerLeuThrPheThrGlyTyrAlaLys 394
 DB 1479 GGTATATGACACCATGAGTGGCAGGTTTATTCATCTATGGGATACGCGGTATGCAAAA 1538
 QY 395 IleHisIleProIleAlaSerValSerGluHisGlnProThrThrTyrValSerPhe 414
 DB 1539 ATACACATTCCAATATTGTCATCAGTGTCTGAACATCAGCCTCAGCATGGGTGCTTTC 1598
 QY 415 PhePheAspLeuHisIleLeuValCysThrPheProIleGlyLeuThrPheCysIleLys 434
 DB 1599 TTCCTTTGATCTACATATTCTTGTATGTACCTTCCCGAGCGCTATGGTTCTGCATCAA 1658
 QY 435 AsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAla 454
 DB 1659 AATATCAACGATGAAGAGTATTGTGCTCTGTATCGATCAGTGTCTACTTTGCC 1718
 QY 455 GlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIle 474
 DB 1719 GGAGTGATGGTGGCGCTGATGCTGACTCTGACCCCGGTCTGTGATGCTGCTCGGCCATC 1778
 QY 475 AlaPheSerAsnValPheGluHisTyrLeuGlyAspMetIleValArgGluAsnProPro 494
 DB 1779 GCCTTCTCAATGTTTTTGGACACTATTTTGGGGATGACATGAAGAAAGGAAACCCACCT 1838
 QY 495 ValGluAspSerAspGluAspAspIleArgAsnGlnGlyAsnLeuTyrAspLysAl 514
 DB 1839 GTGGAGGACGAGTGTGAGGATGACAAAGAAACCC-AGGAACTTGTATGACAAAGC 1897
 QY 514 aGlyLysValArgLysHisAla-ThrGluGlnGluLysThr-GluGluGlyLeuGlyPro 533
 DB 1898 AGGTAAGTGAAGAGCATGT-GACAGAGCAAGAGAA-ACCTGAAGAGGCTTGGGCCCC 1955
 QY 534 AsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHis 553
 DB 1956 AACATCAAAAGCAATGTGACCATGTGTATGCTCTGATGATGTTCGCGGTCAC 2015

QY 554 CysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn 573
 DB 2016 TGCACGTGGGTTCACAGCAACGCTACTCCAGTCCCAAGTGTGCTTGGCTCTCTACAT 2075
 QY 574 HisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGln 593
 DB 2076 CATGATGGTACCAAGGATATATTAGATGATTTTAGAAGACGCTACTTTTGGCTGAGCAA 2135
 QY 594 AsnThrAspGluHisAlaArgValMetSerTyrTrpAspTyrGlyTyrGlnIleAlaGly 613
 DB 2136 AACACGATGAACACCGCCCGGTCTGTGTGGTGGACTACGGCTACAGATTCTGGC 2195
 QY 614 MetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeu 633
 DB 2196 ATGCCCAACAGGACCACTCTGGTGGATAACAACACCTGGAAACAGCCACATCCACTG 2255
 QY 634 ValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr-LeuAs 653
 DB 2256 GTCGNAAGCTATGTCTTCCATGAACCGCCCTATAAATCATGAGTCT-CCTTGA 2314
 QY 653 PValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAs 673
 DB 2315 TGTCCGATTTATGTGTGTTATTTCCGAGGAGTGTGCTATTCGGGGAGCATATCAA 2374
 QY 673 nLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSe 693
 DB 2375 CAAGTTCCTCTGGATGTCAGGATAGCTGAAGGGGAGCATCCCAAAGACATCCGGGAAG 2434
 QY 693 r-AspTyrPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThrLeu 712
 DB 2435 -TGACTATTTTCAC-CCAGCAGGAGAGTTCGAGTAGACAAAGCTGGGTCTCTACTCTG 2492
 QY 713 LeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPhe 732
 DB 2493 TTAAACTGCCTTATGTATAAATGTCTACTACAGATTTGGAGAAATGAGCTAGATTTT 2552
 QY 733 ArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLys 752
 DB 2553 CGCACTCCCCAGCTTTGACCAACACGTAATGCTGAGATTGGAAATTAAGACATTAA 2612
 QY 753 PheLysHisLeuGluGluAlaPheThrSerGluHisThrLeuValArgIleTyrLysVal 772
 DB 2613 TTCAAGCATTTGGAGGAAGCTTTTACATCAGAGCACTGGCTTGTGAGGATATATAAGTG 2672
 QY 773 LysAlaProAspAsnArgGluThrLeu 781
 DB 2673 AAAGCACTGACAAACAGGGAGACACTA 2699
 RESULT 4
 LOCUS CQ782554 1664 bp DNA linear PAT 17-MAR-2004
 DEFINITION Sequence 2694 from Patent EP1396543.
 ACCESSION CQ782554
 VERSION CQ782554.1 GI:45502338
 KEYWORDS Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
 Koga, H.
 TITLE Primers for synthesizing full length cDNA clones and their use
 JOURNAL Patent: EP 1396543-A 2694 10-MAR-2004;
 Research Association for Biotechnology (JP)
 FEATURES
 source 1..1664
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 28..1362
 /note="unnamed protein product"
 CDS

BC052433.1 GI:30851501
MGC.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4236)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Wax, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Sheets, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.D., Hu, X., S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smalusz, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A. 2002.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2388257
12477932
2 (bases 1 to 4236)
Straussberg, R.
Direct Submission
Submitted (15-MAY-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas L. Casavant.
Web site: <http://genome.uiowa.edu>
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,
Fisher, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
Scheetz, T., Smith, C., Snit, E., Tack, D., Trout, K., Walters, J.,
Casavant, T., Soares, M.B.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
1..4236
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="MGC:64679 IMAGE:6837097"
/tissue_type="Brain, enriched mouse brain 12.5dpc"
/clone_lib="NIF BMAP_F10"
/lab_host="DH10B"
/note="Vector: pYX-ASC"
1..4236
/gene="1300006C19rik"
/note="synonym: Simp"
/db_xref="LocusID:68292"

QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
Db 421 ACTGTTTACCAGGGTGTATGATAACCCGCTGGCTTATTCATTGGATTTTAAATACATTG 480
QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
Db 481 AACATAACTGTTTACATAAGAGAGCGTATGTGTGTTTCCCTTGCACCAACTTTTAGCGGCTT 540
QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
Db 541 ACATCTATATACATCTTCTCTGCTTACAAGAGAACTTTTGAACCAAGAGAGAGACTTTTA 600
QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
Db 601 GCTGCTGTGTTTATTGCTATTGTACAGGCTACATATCTCGTTCAGTAGCTGGATCCTTT 660
QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleSer 240
Db 661 GATAATGAAGGCAATGCTATTTTGCACCTTCAGTTTCACATACTATTTATGGTAAATCT 720
QY 241 ValIleThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetVal 260
Db 721 GTAAAAAATCGGCTGAGTCTTTTGGACAATGTGCTGCTTATCTTATTTCTATATGTC 780
QY 261 SerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
Db 781 TCTGCTGGGGTGGTATGATTTATCATCAATCTTATCCACTGCAATGATTTTGGTTG 840
QY 281 LeuLeuMetGlnArgTyrSerIleValTyrIleAlaTyrSerThrPheTyrIleVal 300
Db 841 TTACTGATGCAGATACAGCAAAAGAGCTACATAGCATATAGCACTTTCTACATGTC 900
QY 301 GlyLeuIleLeuSerMetClnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
Db 901 GGTATTAATATCAATGAGATACCTTTTGGGATTTCCAGCAATCAAGAAAGTAA 960
QY 321 HisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
Db 961 CACATGGCAGCTGCAGGTGCTTTGCAATGTCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1020
QY 341 ArgAspArgLeuThrIleGlnIlePheGlnThrLeuPheLeuGlyValSerLeuAla 360
Db 1021 AGAGACCGATTAAACAAACAGAGTCCAGACCTTTTCTTTTGGGTGATCAGTACGT 1080
QY 361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrp 380
Db 1081 GCAGGTGCTGCTTCTTCTAGTGTATCTATTGCTTATACAGTTACATTCACATGTC 1140
QY 381 SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaIleHisIleProIleIle 400
Db 1141 AGTGGCAGGTTTATTCATTGTTGGGATCTGGGTATGCAAAATPACATTCACATTAAT 1200
QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIle 420
Db 1201 GCATCAGTGTCTGAGCATCAACTACGACTTGGGTGCTTCTTCTTGTATCTACATATT 1260
QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLeuAsnIleAsnAspGluArg 440
Db 1261 CTTGTATGACCTTCCAGCAGGCTTTGGTTCTGCATCAAAATATCAACGATGAAGA 1320
QY 441 ValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeu 460
Db 1321 GTATTGTTGCTCTATATGCAATCAGTCTGCTTCTTCTTGTGAGTGTGTTGCGACTG 1380
QY 461 MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
Db 1381 ATGTTGACCTTGCATCCAGTGTGTGTATGCTGTCTGCAATGCTTTTCAATATGTTTT 1440
QY 481 GluHisTyrLeuGlyAspAspValIleArgGluAsnProProValGluAspSerSerAsp 500
Db 1441 GAGCACTATTGTTGGGGATGACATGAAAAAGGAAAAATCCACTGTGGAGGACAGCAGTAT 1500
QY 501 GluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHis 520

Db 1501 GAGGATGCAAAAAGAAACCAAGGAAATTTGTATGTAAGCGAGGTAAAGTGGGAAACAT 1560
QY 521 AlaThrGluGlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleValThr 540
Db 1561 GCAACTGAACAGGAAAAAACTGAAGAGGGATTAGGCCCTTAATATAAAAAGCATTTGTCCAC 1620
QY 541 MetLeuMetLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
Db 1621 ATGTTGATGCTGATGCTATTGATGATGTTTGTCTGCTCACCTGCTACCTGGGTCAAGCAAT 1680
QY 561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
Db 1681 GCCTACTCTAGTCCAGGTAGTCTCTGGCCTCATACAATCATGATGCCACCAAGGAATATC 1740
QY 581 LeuAspAspPheArgGluAlaTyrPheTrpLeuAsnGlnAsnThrAspGluHisAlaArg 600
Db 1741 TTAGATGATTTTAGAGAGCTTTACTTTTGGCTAAGCAAAATACAGATGAACATGACGA 1800
QY 601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeu 620
Db 1801 GTAAATGCTTGGTGGGATTTATGCTATCAGATAGCTGGAATGCTTAATAGAACTACGTTG 1860
QY 621 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
Db 1861 GTGATATATACACTCGAATATACGCCACATAGCACTGTGTGGAAAAAGCTATGCTTCT 1920
QY 641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIle 660
Db 1921 AATGAAACAGCAGCCTATAAAATCATGAGACTCTAGATGTAGATTATCTTTGTTGTTAT 1980
QY 661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
Db 1981 TTTGAGAGGTTATTTGCTATTTCTGGTGATGATATCACAAAATTTCTCTGGATGTTAG 2040
QY 681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
Db 2041 ATAGCTGAAGGAGAAACATCCAAAGACATTCGGGAAAGTGACTATTTTACCCACAGGGA 2100
QY 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
Db 2101 GAATTCGTTGATGACAAAGCAGGATCCCTTCTTGTGAATTCCTTTATGATATAAATG 2160
QY 721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArg 740
Db 2161 TCATACTACAGATTGGAGAAATGCAGTGGATTTTCGTACACCCCGAGGTTTGCACGA 2220
QY 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisIleGluAlaPhe 760
Db 2221 ACAGTAAATGCTGAGATTGGAATTAAGGACATTAATTAATTCACATTTGGAAGAGCTTT 2280
QY 761 ThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
Db 2281 ACATCAGAACACTGGCTTGTAGGATATATAAAGTAAAGACCTGATTAACAGGGAGACA 2340
QY 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
Db 2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTTCARAAGAG 2400
QY 801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
Db 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAA 2460
QY 821 IleSerLysLysThrVal 826
Db 2461 ATATCTAAGAGACTGTT 2478

RESULT 3

BC052433

LOCUS

DEFINITION

IMAGE:6837097), complete cds.

ACCESSION

BC052433 4236 bp mRNA linear ROD 12-NOV-2003
Mus musculus RIKEN cDNA 1300006C19 gene, mRNA (cdna clone MGC:64679
IMAGE:6837097), complete cds.
BC052433

QY 601 ValMetSerTptPAsPtyrGlyTyrGlnIleAlaGlyMetAlaSerArgThrLeu 620
 DB 1801 GTAATGCTCTGTGGCATATGCGTATCAGATAGCTGGAATGCTAATAGACTAGCTTG 1860
 QY 621 ValAspAsnThrTptAsnAsnSerHisIleAlaLeuValGlyCysAlaMetSerSer 640
 DB 1861 GTGGATATATACACTGCTGAATTAACAGCCATAGACTGCTGGGAAAGCTATGCTTCT 1920
 QY 641 AsnGluThrAlaAlaTyrIleMetArgThrLeuAspValAspTyrValLeuValIle 660
 DB 1921 AATGAACAGCAGCCTATATAATCATGAGACTCTAGATCTAGATTATGTTTGGTTATT 1980
 QY 661 PheGlyGlyValIleGlyTyrSerClyAspAspIleAsnLysPheLeuTptMetValArg 680
 DB 1981 TTGGAGGGTATTGGCTATCTGCTGATGATATCAACAAATTCCTCGATGGTTAGG 2040
 QY 681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
 DB 2041 ATAGCTGAAGGAGAAACATCCCAACACATTCGGGAAAGTGACTATTTTACCCACAGGGA 2100
 QY 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
 DB 2101 GAATTCCTGCTAGACAAAGCAGGATCCCTACTTTGTTGATTCCTTATGTAATAATG 2160
 QY 721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArg 740
 DB 2161 TCATACCTACAGATTGGGAGAAATGCGCTGGATTTTCGTACACCCCGAGTTTTCACCGA 2220
 QY 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPhe 760
 DB 2221 ACAGTAACTGCTAGATTGGAAATAGGACATTAATTAACATTTGGAGAGAGCTTT 2280
 QY 761 ThrSerGluHisTptLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
 DB 2281 ACATCAGAACACTGCTGTTAGGATATATAAGTAAAGCACCCTGATACAGGAGACA 2340
 QY 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
 DB 2341 TTAGATATCAACACTCGAGTACCACATTTTCCCAACAGAGATTTTGTCAAGAGAG 2400
 QY 801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
 DB 2401 ACTACCAAGAGAGCGTGGCTACATTAAAAATAAGCTGTTTAAAGAAAGCAAGAAA 2460
 QY 821 IleSerLysLysThrVal 826
 DB 2461 ATATCTAAGAGACGTGTT 2478
 RESULT 2
 AY074880
 LOCUS Homo sapiens source of immunodominant MHC-associated peptides
 DEFINITION (SIMP) mRNA, complete cds.
 ACCESSION AY074880
 VERSION AY074880.1 GI:19879588
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 2481)
 Mc Bride, K., Baron, C., Picard, S., Martin, S., Boismenu, D., Bell, A.,
 Bergeron, J., and Perreault, C.
 TITLE The model B2dmi minor histocompatibility antigen is encoded by a
 mouse homolog of the yeast SPT3 gene
 JOURNAL Immunogenetics 54 (8), 562-569 (2002)
 MEDLINE 22326278
 PUBMED 12439619
 REFERENCE 2 (bases 1 to 2481)
 Mc Bride, K. and Perreault, C.
 AUTHORS
 DIRECT SUBMISSION
 TITLE Submitted (22-JAN-2002) Molecular Biology Group, Compaticigene, 6100
 JOURNAL Royalmount, Montreal, Qc H4P 2R2, Canada

FEATURES
source

gene

CDS

Location/Qualifiers

1. 2481
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 1. 2481
 /gene="SIMP"
 /gene="SIMP"
 /note="similar to Saccharomyces cerevisiae OSTT3 and Mus
 musculus protein encoded by GenBank Accession Number
 AK018758; source of minor histocompatibility antigen(s)"
 /codon_start=1
 /product="source of immunodominant MHC-associated
 peptides"
 /protein_id="AAL71884.1"
 /db_xref="GI:19879588"
 /translation="MAEPAPAPESKHSLSNPSMGLMALNSRHGHGPGACAHKA
 AGAAPKAPAPAGLSGQAPAGWSLISFTILFLAWLAFSSRLFAVIRPESIHET
 DPWVRDTHHLASHGFEFLNWFDERAWYPLGRIVGTVVPGMLITAGLHILNLT
 NITVIRDYCVFLATFESGLTSTSTELTRELNOGAGLAAFCIAIVPGYISRSVAG
 SPNEGIALFALQFTYLYLWVSKVTGSVFWMCCCLSYFVXVSAGGVFIINLPLH
 VFLLLMQRYSKRVIAIYSTFYVGLIILSMQIPFVGFQPIRTSEHMAAGVFAALQAY
 AFLQURDLTKQEFOTFFLGLVLAAGAVFLSVIYLYTGYIAPWSGRFVSLMDTGY
 AKTHIPTIASVSEHPTTWSFFDLHLVCTFPAGLWFCIKNINDEKVFVLAIVSA
 VYPAGVYVRLMLTLTPVVCMLSAIAFSNVFHYGLDDMKRENPPVEDSDDEKRNQ
 NLYDRAGKVRKHATEQETEGLEGNPKISIVTMLMLLMLFAVHTWTSNAYSSVS
 VVLASVNDGTRNILDFFREAYFLRQNTDEHARVMSWMDYGYQTAGMANETTLVDNN
 TWNSHIALVKGAMSNETAAYKIMRTLDVLYFGVYIGYSGDDINKELPMVRIA
 EGHKPDJESDYFTQGEFRVDKSPSTLLNCLMVKMSYRFGEMOLDERTPTPGFDR
 TRNAELGNDIKFKLEAFTSEHNLVLYIKVKAPDNRETLIDHKEPVNTNIPFKQKYL
 KKTTRKRGYIKNKLVFKKGRKSKTKV"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2481
 Scores: 826 00 Matches: 826
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-10-028-384-2 (1-826) x AY074880 (1-2481)

QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTpt 20
 DB 1 ATGCGGAGCCCTCGGCCCGGAGACAGCAAGTCGTCCTCACTCGTCCCGCTGG 60
 QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisHisGlyProGlyAlaGlnCys 40
 DB 61 AGTGGCTCATGCGCCCTGGGAAACAGCGGACGCGCACACGCGGCCCGGGCCAGTGC 120
 QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 DB 121 CGCACAGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 180
 QY 61 GlyLeuSerGlnProAlaGlyTptGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 DB 181 GGGCTGTGCGAGCGGCTGGTGGCAGTCGCTCTCTCCCTTCCACCATCTCTCTCTG 240
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100
 DB 241 TGGCTTCCCGGCTTACCTCGGCCCTCTTCGCGCGTCATCCGCTTCCGAAAGCATCATCC 300
 QY 101 GluPheAspProTptPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyr 120
 DB 301 GAGTTCGACCCGCTGGTTTAACTATAGATCAACACATCATCTTCGATCTCATGGTTCAT 360
 QY 121 GluPheLeuAsnTptPheAspGluAlaTptTptProLeuGlyArgIleValGlyGly 140
 DB 361 GAATTTTAAATTTGGTTTGTATGAAGAGGATGGTATCCACTAGGAAGATACTAGTGTGT 420

AGGAAPKAPAGLGGSLQSPAGWQSLSTFLFLAWLAGFSRSLFAVIRPESIIHF
 DPWNYRSTHLSHGYEFLNFWDERAWPLGRIVGGTVYGLMITAGLIHMIINTL
 NITVHDVCFVLAFTSGLTSTPLTRELWNOGAGLAACTAI VPGYISRVAG
 SPFDNGTAIPALOPTYVLWVKSTGVSFWTCCCLSYFVWVSANGVYFIINLPLH
 VFLVLLMORYSKYVYIAYSTFYVGLLSMOIIPFVGPQIRSEHMAAGAFVALLQY
 AFLQIDRLITKQBFQTLPLFGLVSLAAGAVLSVILYITGHIAPWSGRFYSLWDTGY
 AKIHPIIASVSESPQTTWVSFFDFDHLVCTPFPAGLWFCINRINDERVFVAYIAIS
 VYFAGVRLMLTLTPVVCMLSAIAFNSVFEHYLGDMMKREPPVEDSDEDDKNQ
 NLYDKAGVKRLHATQKTEEGELGPNIKSIVTLMMLLMFMFAVHCITWTSNAYSPPS
 VVLASVHGDTRNITLDDFREAYFWLQNTDEHARVMSWDYGVQAGNARVTLVDNN
 TWNSHIALVGAKMSNETAAKIMRLDVLVYLVIFGVYIGYSGDDINKPLMWYIA
 EGRHPKIDRESDYTPPOGEFRVDKAGSPULLNCLMYKMSVYRFGEMOLDPFRPGFDR
 TRNAEIGNKDIKFKHSEAFSTSHLWRLVYKVPDRNRETLDHKERVYNI FPKYKLS
 KKTTRKRGVINKVLVFKKKSIVKTV"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2481
 Score: 826.00 Matches: 826
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 6

US-10-028-384-2 (1-826) x AX799082 (1-2481)

Qy	1	MetAlaGluProSerAlaProGluSerIleHisLysSerSerLeuAsnSerSerProTrp	20
Db	1	ATGGCGGAGCCCTCGGCGCGGAGGAGCAACACAGTCTGTCCTCAACTCGTCCCGCTGG	60
Qy	21	SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys	40
Db	61	ATGGCCCTCATGGCCCTGGGAAACAGCGCGCACCGCCAGCCCGCGGCGCCAGTGC	120
Qy	41	AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly	60
Db	121	GCGCACAGGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	180
Qy	61	GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla	80
Db	181	GCGCTGTGCGACCGCGCTGGGTGGCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	240
Qy	81	TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis	100
Db	241	TGGCTTGGCGGCTTCAGCTCGGCTCTTCGCGCTCATCGCTTCGAAAGCATCATCCAC	300
Qy	101	GluPheAspProTrpPheAsnTrpArgSerThrHisIleLeuAsnHisGlyPheTrp	120
Db	301	GAGTTTCGACCGCTGGTTTAACTATAGATCAACACATCATCTTCATCTCATGGGTTCTAT	360
Qy	121	GluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGlyGly	140
Db	361	GAAATTTTAAATGGTTTGTATGAAGAGCATGGTATCCATAGGAAGATAGTAGTGGT	420
Qy	141	ThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu	160
Db	421	ACTGTTTACCAGGCTTGATGATACCGCTGGCTTATTCATTTGATTTTAAATACATTCG	480
Qy	161	AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu	180
Db	481	AACATACTGTTCATACATAGAGAGATGTTGTGTCTCTGACCAACTTTTACGGGCTT	540
Qy	181	ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu	200
Db	541	ACATCTATATCTTCTCTCTTACAGAGAACTTTGGAACCAAGGAGGAGGAGCTTTTA	600
Qy	201	AlaAlaCysPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySerPhe	220
Db	601	GCTGCTGTTTATTTGCTATGTACAGGCTACATATCTCGGTCAGTAGCTGATCTT	660
Qy	221	AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpValLysSer	240
Db	661	GATAATGAAGGCATTTGCTATTTTTCACCTTCAGTTCATACATCTTATTTATGGGTAAATCT	720

Qy	241	VallysThrGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMetVal	260
Db	721	GTAATAAATCGGTCAGCTTTTGGGCAATGTGCTGCTGCTTATCTTATATATGCTC	780
Qy	261	SerAlaTrpGlyGlyTrpValPheIleIleAsnLeuIleProLeuHisValPheValLeu	280
Db	781	TCGCTTGGGCTGGTATGATGATTTATCAATCTTATTCCTACTGATATTTGTGTG	840
Qy	281	LeuLeuMetGlnArgTrpSerIleArgValTrpIleAlaTrpSerThrPheTrpIleVal	300
Db	841	TTACTGATGACAGATACACGAAAGAGCTCATAGCATATAGCACCTTCTACATTTGTG	900
Qy	301	GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu	320
Db	901	GGTTTAAATATCAATGACAGATACCTTTTGGGGAATCCAGCAATCAGAAAGATGAA	960
Qy	321	HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeu	340
Db	961	CACATGGCAGCTGACAGTGTCTTTGCAATCTTATTCCTACTGATATTTGTGATCTG	1020
Qy	341	ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla	360
Db	1021	AGAGCCGATTAACAAACCAAGAGTCCAGACCCCTTTCTTTGGGTGTATCACTAGCT	1080
Qy	361	AlaGlyAlaValPheLeuSerValIleTrpLeuThrTrpGlyTrpIleAlaProTrp	380
Db	1081	GCAGGTGCTGTCTTCTTAGTGTCTATTTGACTTATACAGGTATACATTGCACCATGG	1140
Qy	381	SerGlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaIleHisIleProIleIle	400
Db	1141	AGTGGCAGGTTTTTATCTTGTGGGATCTGGGTATGCAAAATACACATTCATTTAT	1200
Qy	401	AlaSerValSerGluHisGlnProThrTrpValSerPhePhePheAspLeuHisIle	420
Db	1201	GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGTCTTCTTCTTCTCATCATATT	1260
Qy	421	LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg	440
Db	1261	CTTGATGTACCTTCCAGCAGGCTTGTGTCTGTCATCAAAATATCAACATGAAAGA	1320
Qy	441	ValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetValArgLeu	460
Db	1321	GTAATTTGCTCTATATGATCAATCAGTGTCTCTACTTTGCTGGAGTATGCTGGACTG	1380
Qy	461	MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe	480
Db	1381	ATGTTGACTTTGACTCCAGTCGTGTATGCTGCTGCAATTCCTTTTCAAAATGTTTT	1440
Qy	481	GluHisTrpLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSerAsp	500
Db	1441	GAGCCTATTTGGGGGATGACATGAAAGGGAATCCACCTGTGGAGCAGCAGTGTAT	1500
Qy	501	GluAspAspLysArgAsnGlnGlyAsnLeuTrpAspLysAlaGlyLysValArgLysHis	520
Db	1501	GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGAGGTTAAAGTAGGAAACAT	1560
Qy	521	AlaThrGluGlnGluLysThrGluGlyLeuGlyProAsnIleLysSerIleValThr	540
Db	1561	GCAACTGACAGGAAACAACTGAAGAGGGATTAAGGCCCTTAATATAAAAGCATTTGTCAC	1620
Qy	541	MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn	560
Db	1621	ATGTTGATGCTGATGCTTATGATGATGTTTGTGCTGCTCTACTGTACCTGGGTCAAGCAAT	1680
Qy	561	AlaTrpSerSerProSerValValLeuAlaSerTrpAsnHisAspGlyThrArgAsnIle	580
Db	1681	GCCTACTCTAGTCCAAAGTGTAGTCTGGCTCATCAATCATGATGGCAGCAGGAATATC	1740
Qy	581	LeuAspAspPheArgGluAlaTrpPheTrpLeuArgGlnAsnThrAspGluHisAlaArg	600
Db	1741	TTAGATGATTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACGA	1800

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 9151.03 Seconds
(without alignments)
4268.513 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 826

Sequence: 1 MAEPSAPSKHSSNSPW.....GYKNKLVPFKGKKISKTV 826

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-O=/cgn2.1/USPFO/spool/US10028384/runat_14122004_131645_6800/app_query.fasta.1.3740
-DB=GenEmbl -QMT=fscap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN 1 1 20262 @runat_14122004_131645_6800 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

Result No.	Score	Match	Length	DB	ID	Description
1	826	100.0	2481	6	AX799082	AX799082 Sequence
2	826	100.0	2481	9	AY074880	AY074880 Homo sapi
3	452	54.7	4236	10	BC052433	BC052433 Mus muscu
4	442	53.5	1664	6	CQ782554	CQ782554 Sequence

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	826	100.0	2481	6	AX799082	AX799082 Sequence
2	826	100.0	2481	9	AY074880	AY074880 Homo sapi
3	452	54.7	4236	10	BC052433	BC052433 Mus muscu
4	442	53.5	1664	6	CQ782554	CQ782554 Sequence

RESULT 1

AX799082
LOCUS AX799082 2481 bp mRNA linear PAT 08-OCT-2003
DEFINITION Sequence 1 from Patent WO03054008.
ACCESSION AX799082
VERSION AX799082.1 GI:37605057
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Perreault, C. and McBride, K.
Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy
Patent: WO 03054008-A 1 03-JUL-2003;
Compatisgene Inc. (CA)
FEATURES
source
1..2481
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..2481
/codon_start=1
/product="SIMP"
/protein_id="CAE48447.1"
/db_xref="GI:37605058"
/translation="MAEPSAPSKHSSNSPWGLMALGNLSRHGHGPGCAQCAHKA"

ALIGNMENTS

5	442	53.5	1664	6	BD127193	Primer fo
6	442	53.5	1664	9	AK074587	AK074587 Homo sapi
7	365	44.2	2710	6	AX799084	AX799084 Sequence
8	323	39.1	2546	6	AX099510	AX099510 Sequence
9	323	39.1	2546	6	BD063986	BD063986 Secreted
10	317	38.4	2510	6	AX136135	AX136135 Sequence
11	317	38.4	2510	6	BD123520	BD123520 Secretary
12	317	38.4	2510	9	AK075380	AK075380 Homo sapi
13	289	35.0	2503	6	AX017997	AX017997 Sequence
14	289	35.0	2503	6	BD137303	BD137303 Human nuc
15	286	34.6	2284	6	AX882932	AX882932 Sequence
16	286	34.6	2284	6	BD160013	BD160013 Primer fo
17	286	34.6	2284	9	AK027789	AK027789 Homo sapi
18	246	29.8	957	9	BC015880	BC015880 Homo sapi
19	195	23.6	1734	10	BC003206	BC003206 Mus muscu
20	168	20.3	507	6	AX886102	AX886102 Sequence
21	168	20.3	507	6	BD025712	BD025712 Sequence
22	167	20.2	787	6	AX869456	AX869456 Sequence
23	167	20.2	787	6	BD149518	BD149518 Primer fo
24	167	20.2	1040	6	CQ723424	CQ723424 Sequence
25	145	17.6	784	6	AX136480	AX136480 Sequence
26	145	17.6	784	6	BD123720	BD123720 Secretary
27	143	17.3	433	6	AX340470	AX340470 Sequence
28	133	16.1	558	6	AX079689	AX079689 Sequence
29	123	14.9	2508	10	BC013054	BC013054 Mus muscu
30	117	14.2	616	6	BD210442	BD210442 Human gen
31	114	13.8	2620	5	CR386955	CR386955 Gallus ga
32	107	13.0	349	6	BD071886	BD071886 Secreted
33	105	12.7	170691	9	AC104643	AC104643 Homo sapi
34	97	11.7	101046	2	AC117975	AC117975 Rattus no
35	97	11.7	203492	2	AC107826	AC107826 Mus muscu
36	97	11.7	227953	2	AC134292	AC134292 Rattus no
37	97	11.7	285379	2	AC113070	AC113070 Mus muscu
38	82	9.9	250	6	AX885575	AX885575 Sequence
39	82	9.9	250	6	BD025185	BD025185 Sequence
40	81	9.8	49784	2	AC101107	AC101107 Mus muscu
41	78	9.4	245	6	AX905227	AX905227 Sequence
42	78	9.4	245	6	BD040760	BD040760 Sequence
43	78	9.4	247	6	BD075907	BD075907 5' EST of
44	72	8.7	177253	9	AC092024	AC092024 Homo sapi
45	63	7.6	1748	5	BC052114	BC052114 Danio rer

```

Db 303 AATGAAACAGCAGCCTATATAAATCATGAGGACTCTAGATCTAGATTATGTTGGTTATT 362
Qy PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetValArg 680
Db 363 TTTCGAGGGTTATTTGGCTATTCGGTGATGATATCAACAAATTTCTCTGATGGTTAGG 422
Qy IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
Db 423 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTCACTATTTACCCCCACAGGGA 482
Qy GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
Db 483 GAATTCCTGTAGACAAAGCAGGATCCCCCTACTTTGTGAATTCCTTATGTATAAAATG 542
Qy SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArg 740
Db 543 TCATACTACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCCCAGGTTTGGACCGA 602
Qy ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPhe 760
Db 603 ACACGTAAATCTGAGATTGGAAATAAGGACATTAAATTCAAACATTTGGAAGAAGCCTTT 662
Qy ThrSerGluHisTyrLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
Db 663 ACATCAGACACTGGCTTGTAGATATATAAAGTAAAGCCCTGATTAACAGGGAGACA 722
Qy LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
Db 723 TTATGTCACAAACCTCGAGTCACCAACATTTCCCAAAACAGAAATATTTGTCAAAGAAG 782
Qy ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
Db 783 ACTACCAAAAGGAAGCGTGGCTACATTAATAATTAAGCTGGTTTTTAAGAAAGGCAAGAA 842
Qy IleSerLysLysThrVal 826
Db 843 ATATCTAAGAAGACTGTT 860

```

Search completed: December 14, 2004, 23:34:06
 Job time : 988.69 secs

```

PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX
XX Claim 8; SEQ ID NO 17837; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dr primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX oligonucleotides, all of which are used in the exemplification of the
XX present invention
XX
XX SQ Sequence 2284 BP; 726 A; 388 C; 436 G; 734 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8,12e-284 Length: 2284
Score: 286.00 Matches: 286
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 34.62% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x AAH18021 (1-2284)
QY 541 MetLeuMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
DB 3 ATGTTGATGCTGATGCTATTGATGATGTTTCTGCTCCACGTACCTGGGGTCAAGCAAT 62
QY 561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
DB 63 GCCTACTCTAGTCCAAAGTGTAGTCTTGGCTCATACATCATGATGCGACCAATATC 122
QY 581 LeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
DB 123 TTAGATGATTTAGAGAAGCTTACTTTGGCTTAAGGCAAAATACAGATGACATGACGCA 182
QY 601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
DB 183 GTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGAAATGCTTAATAGAACTACGTTG 242
QY 621 ValAspAsnAenThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
DB 243 GTGGATATAACACCTCGAATAACAGCCACATAGCACTGGTGGGAAAGCTATGCTTCT 302
QY 641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValValLeuValIle 660
DB

RESULT 15
AAH18021
ID AAH18021 standard; cDNA; 2284 BP.
XX
XX AC AAH18021;
XX
XX 26-JUN-2001 (first entry)
XX
XX DE Human cDNA sequence SEQ ID NO:17837.
XX
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX OS Homo sapiens.
XX
XX PN EPI074617-A2.
XX
XX PD 07-FEB-2001.
XX
XX EF 28-JUL-2000; 2000EP-00116126.
XX
XX PR 29-JUL-1999; 99JP-00248036.
XX
XX PR 27-AUG-1999; 99JP-00300253.
XX
XX PR 11-JAN-2000; 2000JP-00118776.
XX
XX PR 02-MAY-2000; 2000JP-00183767.

```

```

Db 361 CACATAGCAGCTGGTGGGAAAAAGCTATGCTTCTTAATGAAACAGCAGCCTATAAATCATG 420
Qy 650 ArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 669
Db 421 AGGACTAGATGATAGATATGTTTGGTATATTTTGGAGGGGTATGCTATTCCTGGT 480
Qy 670 AspAspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAsp 689
Db 481 GATGATATCAACAAATTTCTCTGGATGTTAGGATAGCTGAAGGAGAACATCCCAAGAC 540
Qy 690 IleArgGluSerAspTyrPheThrProGluGlyGluPheArgValAspLysAlaGlySer 709
Db 541 ATTGGGGAAGTGAATTTTACCCACAGGAGAAATTCCTGTAGACAAAGCAGATCC 600
Qy 710 ProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGln 729
Db 601 CCTACTTTGTTGATTCCTTATGTATATAAATGTCATACATACAGATTTGAGAAATGCAG 660
Qy 730 LeuAspPheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLys 749
Db 661 CTGATTTTCGTACACCCCGGCTTTTGCACGACAGCTAATGCTGAGATTCGAAATAAG 720
Qy 750 AspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIle 769
Db 721 GACATTAATTCAAACATTTGGAGAGCCTTTACATCAGAACACTGGCTTGTGGATA 780
Qy 770 TyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAsn 789
Db 781 TATAAGTAAGACCCCTGTATAACAGGAGACATTAGATCACAAACCTCGAGTCACCAAC 840
Qy 790 IlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIle 809
Db 841 ATTTCCCAAAACAGAGATTTTGTCAAGAGACATCCAAAGGAGGCTGGCTACATT 900
Qy 810 LysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLysThrVal 826
Db 901 AAAAATAAGCTGTTTAAAGAAAGCAAGAAATATCTAAGAAAGACTGTT 951

RESULT 14
AAD08289 standard; cDNA; 2537 BP.
XX
XX AAD08289;
XX
XX 08-AUG-2001 (first entry)
XX
DE Human secreted protein-encoding gene 7 cDNA clone HDTLR06, SEQ ID NO: 17.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnerability; binding partner identification;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 93..953
FT /tag= a
FT /product= "Human secreted protein precursor"
FT sig_peptide 93..155
FT /tag= b
FT mat_peptide 156..950
FT /tag= c
FT /product= "Mature human secreted protein"
XX
FN WO200136440-A1.

```

```

XX PD 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-US031282.
XX
XX 19-NOV-1999; 99US-0166414P.
XX
XX 21-JUL-2000; 2000US-0219665P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsculis GA, Birse CE, Moore PA;
XX WPI; 2001-343795/36.
XX P-PSDB; AAE03824.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.
XX
XX Claim 1; Page 443; 553pp; English.
XX
XX AAD08283-AAD08355 represent cDNAs corresponding to 23 human secreted
XX protein genes, and AAE03818-AAE03870 represent the proteins they encode.
XX AAE03871-AAE03896 represent human secreted protein fragments or variants.
XX The secreted proteins and their genes are useful for preventing, treating
XX or ameliorating medical conditions, e.g., by protein or gene therapy.
XX Pathological conditions can be diagnosed by determining the amount of the
XX new protein in a sample or by determining the presence of mutations in
XX the new genes. Specific uses are described for each of the 23 genes,
XX based on the tissues in which they are most highly expressed, and include
XX developing products for the diagnosis or treatment of proliferative
XX disorders, cancer, tumours, foetal and developmental abnormalities,
XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
XX angiogenic disorders, kidney disorders, gastrointestinal disorders,
XX pregnancy-related disorders, endocrine disorders, and infections. The
XX proteins can also be used to aid wound healing and epithelial cell
XX proliferation, to prevent skin aging due to sunburn, to maintain organs
XX before transplantation, for supporting cell culture of primary tissues,
XX to regenerate tissues, to identify their cognate ligands or binding
XX partners, and in chemotaxis, and can be used as a food additive or
XX preservative to modify storage properties. Antibodies specific for a
XX protein of the invention can be used in alleviating symptoms associated
XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,
XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
XX present sequence represents a human secreted protein-encoding cDNA of the
XX invention
XX
XX Sequence 2537 BP; 816 A; 434 C; 487 G; 800 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. NO.: 4.2e-295 Length: 2537
XX Score: 297.00 Matches: 297
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 35.96% Indels: 0
XX DB: 4 Gaps: 0
XX
XX US-10-028-384-2 (1-826) x AAD08289 (1-2537)
XX
XX Qy 530 GlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetMet 549
XX Db 60 GGTTAGGCCCTAATATAAAGACATGTCCACCATGTTGATGCTATGATGATG 119
XX
XX Qy 550 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerProSerValValLeu 569
XX Db 120 TTTGCTGTCCACTGTACCTGGGTCAACAGCAATGCTACTCTAGTCCAGTGTAGTCTCG 179
XX
XX Qy 570 AlaserTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPhe 589
XX Db 180 GCCTCATACATCATGATGGCACCAGGAATATCTTAGATGATTTTAGAGAGCTACTTTT 239

```

QY 604 TrpThrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsn 623
 Db 301 TGGTGGGATATGCTATCAGATAGCTGGAAATGGCTAATAGAACTAGCTTGGTGGTAAAT 360
 QY 624 AsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThr 643
 Db 361 AACACCTGGAAATAACAGCCACATAGCACTGGTGGGAAAAGCTATGCTCTTCTAATGAAACA 420
 QY 644 AlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGly 663
 Db 421 CGACGCTATAAATCATAGAGACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGG 480
 QY 664 ValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGlu 683
 Db 481 GTTATTGGCTATTCTGGTGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGTAA 540
 QY 684 GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
 Db 541 GGAGAACATCCCAAGACATTCGGGAAGTGCATATTTACCCACAGGAGAAATTCCT 600
 QY 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
 Db 601 GTAGACAAAGCAGGATCCCTACTTTGTTGAATTCCTTATGTATATAAATGTCACTAC 660
 QY 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn 743
 Db 661 AGATTGGAGAAATCGAGCTGGATTTTCGTACCCCCAGGTTTTCACCGAACACGTAT 720
 QY 744 AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPheThrSerGlu 763
 Db 721 GCTGAGATTGGAATAAGACATTAATTAATTCACCAATTTGGAAGAGCCTTTACATCAGAA 780
 QY 764 HisTrpLeuValAlaGlyTyrLysValLysAlaProAspAsnArgGluThrLeuAspHis 783
 Db 781 CACTGGCTTGTTAGATATATAAGTAAAGCAACCTGATACAGGAGACATTAGATCAC 840
 QY 784 LysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLys 803
 Db 841 AAACCTCGAGTCACCAACATTTCCCAAAACAGAGATATTGTCAAAGAGACTACCAAA 900
 QY 804 ArgLysArgGlyTyrLysLysAsnLysLeuValPheLysLysGlyLysLysLysLys 823
 Db 901 AGGAAGCGTGCTACATTAATAATAGCTGGTTTTTAAGAAAGCAAGAAATATCTAAG 960
 QY 824 LysThrVal 826
 Db 961 AAGACTGTT 969
 RESULT 13
 ID AAF93772 standard; cDNA; 2510 BP.
 XX
 AC AAF93772;
 DT 23-MAY-2001 (first entry)
 XX
 DE Human cDNA encoding a membrane or secretory protein clone PSE0070.
 XX
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1067182-A2.
 XX
 PD 10-JAN-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114090.
 XX
 PR 08-JUL-1999; 99JP-00194179.
 PR 11-JAN-2000; 2000JP-00118775.
 PR 02-MAY-2000; 2000JP-00183766.
 XX

PA (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 XX WPT; 2001-093989/11.
 DR P-PSDB; AAB88345.
 XX
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 XX gene therapy or as candidate target molecules in drug development.
 XX
 PS Claim 1; SEQ ID NO 57; 609pp + Sequence Listing; English.
 XX
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by AAB88317
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
 CC invention. The invention also includes methods for the production of
 CC antibodies directed against the proteins, and cDNA sequences, which can
 CC be used in vaccines. The polynucleotide sequences can be used in gene
 CC therapy. The polynucleotide sequences and the proteins they encode may be
 CC used in the prevention, treatment and diagnosis of diseases associated
 CC with inappropriate secretory protein/membrane protein expression. The
 CC nucleic acids and complementary sequences may also be used as DNA probes
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
 CC and quantitate the presence of similar nucleic acid sequences in samples.
 CC They may also be used to study the expression and function of secretory
 CC proteins/membrane polypeptides and their role in metabolism. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC against them and in assays to identify modulators (agonists and
 CC antagonists) of expression and activity. The antibodies and antagonists
 CC may also be used as therapeutic agents to down regulate expression and
 CC activity. The antibodies may also be used as diagnostic agents for
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme
 CC linked immunosorbent assay (ELISA). Examples of diseases which may be
 CC treated include rheumatoid arthritis and diabetes
 XX
 SQ Sequence 2510 BP; 802 A; 412 C; 489 G; 807 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,06e-315 Length: 2510
 Score: 317.00 Matches: 317
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.38% Indels: 0
 DB: 5 Gaps: 0
 US-10-028-384-2 (1-826) x AAF93772 (1-2510)
 QY 510 LeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnLysThrGluGlu 529
 Db 1 TTGTATGATAGGACGATTAAGTGGAGAAACATGCACTGACAGAAAAAACTGAAGAG 60
 QY 530 GlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMet 549
 Db -61 GGATTAGGCCCTAATAATAAAGACATTTGTCACCATGTTGATGCTGATGCTATGATGATG 120
 QY 550 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeu 569
 Db 121 TTGTGCTGCACGTACCTGGGTGACAGCAATGCTACTCTAGTCCAAAGTGTAGCTCTG 180
 QY 570 AlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPhe 589
 Db 181 GCCTCATCAATCATCATGCGACAGGAGATATCTTAGATGATTTTAGAGAGAGCTTACTTT 240
 QY 590 TrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTyrAspTyrGlyTyr 609
 Db 241 TGGCTAAGGCAAAATACAGATGAACATGACAGGATATGCTTTGTGTGGATTATGGCTAT 300
 QY 610 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 629
 Db 301 CAGATAGCTGGAATGGCTAATAGAACTACCTGGTGGATATAACACCTGGAATAACAGC 360
 QY 630 HisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMet 649

Db 121 CTGATGCTATTGATGATGTTTGCTGCTCACTGTACCTGGGTCAAGCAATGCTACTCT 180
Qy 564 SerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnHisLeuAspAsp 593
Db 181 AGTCCAAGTGTAGTCTGGCTCATACATCATGATGCGACAGGATATCTTAGATGAT 240
Qy 584 PheArgGluAlaTyrPheThrLeuArgGlnAsnThrAspGluHisAlaArgValMetSer 603
Db 241 TTTAGAGAAGCTTACTTTTGCTTAAGCAAAATACAGATGAACATGCACGAGTAATGTCT 300
Qy 604 TrpTrpAspTyrGlyTyrGlnHisAlaGlyMetAlaAsnArgThrThrLeuValAspAsn 623
Db 301 TGGTGGGATATGCTATCATGATAGCTGGAATGGCTAATAGACTACGTTGGGGAAT 360
Qy 624 AsnThrTrpAsnAsnSerHisLeuAlaLeuValGlyLysAlaMetSerSerAsnGluThr 643
Db 361 AACACCTGGGAATACAGCCCATAGCACTGGTGGGAAAGCTATGCTCTTAATGAACA 420
Qy 644 AlaAlaTyrLysLysMetArgThrLeuAspValAspTyrValLeuValLlePheGlyGly 563
Db 421 GCAGCCTATAAATCATGAGACTCTAGATGTAGATTAATGTTTGGTTATTTTGGAGG 480
Qy 564 ValIleGlyTyrSerGlyAspAspLleAsnLysPheLeuTrpMetValArgLleAlaGlu 683
Db 481 GTTATTGGCTATTCTGCTGATGATATCAACAAATTTCTCTGGATGCTTAGGATGCTGAA 540
Qy 684 GlyGluHisProLysAspLleArgGluSerAspTyrPheThrProGlnGlyLlePheArg 703
Db 541 GGAGACATCCCAAGACATTCGGGAAGTGATATTTTACCCCAAGGAGAAATTCCT 600
Qy 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
Db 601 GTAGACAAAGCAGATCCCTACTTTGTTGATTTGCTTATGTATATAAATGTCATAC 660
Qy 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProGlnGlyPheAspArgThrArgAsn 743
Db 661 AGATTGGGAATGAGCTGGATTTTGGTACACCCCAAGGTTTGGACCAAGACGTAAT 720
Qy 744 AlaGluIleGlyAsnLysAspLleLysPheLysHisLeuGluGluAlaPheThrSerGlu 763
Db 721 GCTGAGATTGGAATTAAGACATTAATTCACATTTGGAAGAGCCTTTACATCAGNA 780
Qy 764 HisTrpLeuValArgLleTyrLysValLysAlaProAspAsnArgGluThrLeuAspHis 783
Db 781 CACTGCTGTTAGGATATATAAAGTAAAGCACCTGTATAACAGGAGACATTAGATCAC 840
Qy 784 LysProArgValThrAsnLlePheProLysGlnLysTyrLeuSerLysLysThrThrLys 803
Db 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAGAGACTACCAAA 900
Qy 804 ArgLysArgGlyTyrLleLysAsnLysLeuValPheLysLysGlyLysLysLysSerLys 823
Db 901 AGGAAGCGTGGCTACATTAAATAAATAGCTGGTTTAAAGAAAGGCAAGAAATATCTAAG 960
Qy 824 LysThrVal 826
Db 961 AGACCTGTT 969

RESULT 12

AAF98463 standard; cDNA; 2546 BP.

AC AAF98463;

XX 07-JUN-2001 (first entry)

DE Human cDNA clone CT585_1 sequence SEQ ID 150.

XX Human; secreted protein; nutrient; cytokine modulator; proliferation;
KW differentiation; immune system modulator; tissue growth; chemotactic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss;
KW haematopoiesis.

XX Homo sapiens.
XX WO200119988-A1.
XX PD 22-MAR-2001.
XX 14-SEP-2000; 2000WO-US0251135.
XX 17-SEP-1999; 99US-00398829.
XX (GENY) GENETICS INST INC.
XX Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
PI WPI: 2001-244801/25.
XX P-PSDB; AAB90727.
XX Isolated nucleic acids encoding polypeptides, useful for modulating e.g.
PT cytokine and cell proliferation/differentiation activity, the immune
PT system and hematopoiesis regulating activity.
XX Disclosure; Page 476-477; 557pp; English.
XX Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
CC tissue types, and may be used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate protein expression. The
CC polypeptides and nucleic acids may be used as nutrients or to modulate
CC cytokine and cell proliferation/differentiation activity and may also be
CC involved in modulation of the immune system. The cDNA sequences,
CC proteins, their agonists and/or antagonists exhibit haematopoiesis
CC regulating activity; tissue growth activity; activin/inhibin activity;
CC chemotactic/chemokinetic activity; haemostatic and thrombolytic activity;
CC receptor/ligand activity; anti-inflammatory activity; haematopoiesis
CC activity; cadherin/tumour suppressor activity; and/or tumour inhibition
CC activity. Included in the invention are probes represented in AAF98490 -
CC AAF98572 which are specific for the cDNA clones encoding the secreted
CC proteins
XX
SQ Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0 Length: 2546
Score: 323.00 Matches: 323
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.10% Indels: 0
DB: 5 Gaps: 0

US-10-028-384-2 (1-826) x AAF98463 (1-2546)
Qy 504 LysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGlu 523
Db 1 AAAAGAAACCAAGGAAATTTGTATGATAGGCGAGGTAAGGAGGAAATGCACTGAA 60
Qy 524 GlnGluLysThrGluGlyLeuGlyProAsnLleLysSerLleValThrMetLeuMet 543
Db 61 CAGGAAAACTGAGAGGGAATAGCCCTAATATAAAGCAATTCACCATGTTGATG 120
Qy 544 LeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSer 563
Db 121 CTGATGCTATTGATGATGTTTGTCTGCTCCTGCTACCTGGGTCAAGCAATGCTACTCT 180
Qy 564 SerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnLleLeuAspAsp 583
Db 181 AGTCCAAGTGTAGTCTGGCTCATACATCATGATGCGACAGGATATCTTAGATGAT 240
Qy 584 PheArgGluAlaTyrPheThrLeuArgGlnAsnThrAspGluHisAlaArgValMetSer 603
Db 241 TTTAGAGAAGCTTACTTTTGGCTTAAGCAAAATACAGATGAACATGCACGAGTAATGTCT 300

```

ID 2 TTGCGCGTCATCCGCTTGAAGACATCATCCAGGTTCCGACCGTGGTTAACTATAGA 61
XX
AC
XX
XX 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
XX
XX 62 TCACACATCATCTTGCATCTCATGGTCTCATGAATTTTAAATGGTTGATGAAAGA 121
XX
XX 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
XX
XX 122 GCATGGTATCCACTAGGAAGAAATAGTGGTCTCTGTTTACCCAGGTTGATGATACC 181
XX
XX 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
XX
XX 182 GTGGCCATTATTCATGGATTTTAAATACATTAACATAAATCTTTCACATAAGACGTA 241
XX
XX 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
XX
XX 242 TGTGTGTTCTTCGCCCACTTTTAGCGCCCTTACATCTATATCTACTTCTCTGTTTACA 301
XX
XX 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
XX
XX 302 AGAAGACTTTGGAACCAAGACGAGGACTTTTAGCTGCTGTTTATTGCTATTGTACCA 361
XX
XX 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
XX
XX 362 GGCTACATATCTCGGTCACTAGTGGATCTTTGATAATGAAGGCAATTCCTATTTTGA 421
XX
XX 230 LeuGlnPheThrTyrTrpLeuTrpValLysSerValLysThrGlySerValPheTrpThr 249
XX
XX 422 CTTCACTTACATATTTATTTAGGGTAAATCTGTAAACCTGGTCACTTTTGGACA 481
XX
XX 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
XX
XX 482 ATGTGCTGCTTATCTTCTATATGTTCTGCTGCTGGGTGGTATTGTTTATTC 541
XX
XX 270 IleAsnLeuIleProIleHisValPheValLeuLeuMetGlnArgTyrSerLysArg 289
XX
XX 542 ATCAATCTTATTCATCGATGATGTTTGTGTGTACTGATGAGATACAGCAAGAAGA 601
XX
XX 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
XX
XX 602 GTTACATAGCATATAGCACTTCTACATTTGGGTTTAAATATTATCATGACAGTACT 661
XX
XX 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
XX
XX 662 TTTGTGGGATTCAGGCCAATCAGAACAGTGAACATGCGAGCTGCAGGTGCTTTGCA 721
XX
XX 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPhe 349
XX
XX 722 TTGCTGCAAGCTTATGCTTCTTGCAGTATCTGAGAGACCGATTAAACAAACAAGATTC 781
XX
XX 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
XX
XX 782 CAGACCCCTTTCTTTTGGGTGATCACTAGCTGCAGGTGCTGTGTTCTTGTAGTGTATC 841
XX
XX 370 TyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAsp 389
XX
XX 842 TATTGTGCTTATACAGGTTCATTTGACCAATGAGTGGCAGGTTTATTCATTTGGGAT 901
XX
XX 390 ThrGlyTyrAlaLysIleHisIleProIleAlaSerValSerGluHisGlnProThr 409
XX
XX 902 ACTGGGTATGCAAAATATACATTCCAATATTGTCATCAGTGTCTGAGCATCAACCTAGC 961
XX
XX 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProIleGlyLeu 429
XX
XX 962 ACTTGGGTGCTTCTTCTTGTGATCTACATATTCTTGTGTTGTTACCTTCCAGAGCCCTT 1021
XX
XX 430 TrpPheCysIleLysAsnIleAsnAspGluArg 440
XX
XX 1022 TGGTTCGTCATCAAAATATATCAACGATGAAAGA 1054
XX

```

RESULT 11
AAV44866

AAV44866 standard; cDNA; 2546 BP.

AAV44866;

21-OCT-1998 (first entry)

Clone CT585_1 coding sequence.

Secreted protein; nutritional source; cell proliferation activity; cell differentiation activity; immune stimulant; tissue growth activator; haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor; tumour inhibitor; clone CT585_1; ds.

Homo sapiens.

Key Location/Qualifiers
CDS 112..972
/*tag= a

W09825962-A2.

18-JUN-1998.

12-DEC-1997; 97WO-US023224.

13-DEC-1996; 96US-00766263.

11-DEC-1997; 97US-00989232.

(GEM) GENETICS INST INC.

Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M; Spaulding V, Agostino MG, WPI; 1998-362424/31.
P-PSDB; AAW69247.

New isolated polynucleotides - obtained from human adult testis, human adult ovary, human adult brain and human adult heart cDNA libraries.

Claim 35; Page 79-81; 108pp; English.

This sequence represents a polynucleotide of the invention, and encodes a secreted protein. It was isolated from a human adult brain cDNA library, and is designated clone CT585_1. The DNA sequences and encoded polypeptides can be used as nutritional sources or supplements, or may exhibit e.g. cytokine and cell proliferation/differentiation activity, immune stimulating or suppressing activity, haematopoiesis regulating activity, receptor/ligand activity, anti-inflammatory activity, activin/inhibin activity, chemostatic/chemokinetic activity, cadherin/tumour invasion suppressor activity, tissue growth activity, tumour inhibition activity or other activities

Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	Length:
0	2546
323.00	Matches: 323
Score:	Conservative: 0
Percent Similarity: 100.00%	Mismatches: 0
Best Local Similarity: 100.00%	Indels: 0
Query Match: 39.10%	Gaps: 0
DB:	2

US-10-028-384-2 (1-826) x AAV44866 (1-2546)

QY 504 LysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGlu 523

DB 1 AAAGAACACCAAGGAATTTGATGATAAGGCAGGTAAAGTGAGAAACATGCAACTGAA 60

QY 524 GlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMet 543

DB 61 CAGGAAAAAACTGAAGAGGGGATTAGGCCCTTAATAATAAAAGCATTTGCACATGTTGATG 120

QY 544 LeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSer 563

QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
DB 362 GCCTACATATCTCGTCTAGTGGATCCCTTTCATATAGAGGCATGCTATTTTGA 421
QY 230 LeuGlnPheThrTyrTyrLeuTrpValIleSerValIleThrGlySerValPheTrpThr 249
DB 422 CTTTCAGTTTCACATATCTATTTATGGTAAATCTGTAAACCTGGTGCAGTTTTTTGGACA 481
QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIle 269
DB 482 ATGTGCTGCTGCTTATCTATTTCTATATGCTCTCTGCTGGGTGGTGTATGATTATC 541
QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerLysArg 289
DB 542 ATCAATCTATTTCCACTGCATGATTTTGTGTGTACTGTGATGAGATACACAAAGA 601
QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
DB 602 GTCTACATAGCATATGACATTTCTACATTTGCTGGTTTAAATATATCAATGCAGATACCT 661
QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaIleAlaGlyValPheAla 329
DB 662 TTGTGGGATTCACCCCAATCAGAACAGTGAACATGCGAGCTGCGAGTGTCTTTGCA 721
QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPhe 349
DB 722 TTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGAGACCGATTAAACAAACAGAGTTC 781
QY 350 GlnThrLeuPhePheGlyValSerLeuAlaIleGlyValPheLeuSerValIle 369
DB 782 CAGACCCCTTTCTTTGGGTGATCATCTAGCTCAGGTGCTGTGTTCTTATGTCATC 841
QY 370 TyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAsp 389
DB 842 TATTTGACTTATACAGTATACATTCACCATGGAGTGGCAGGTTTATTCATTTGGGAT 901
QY 390 ThrGlyTyrAlaValIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
DB 902 ACTGGTATGCAAAATACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 961
QY 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
DB 962 ACTTGGGTGCTTTCTTCTTTGATCTACATATCTTGTATGTTACCTTCCAGCAGGCTT 1021
QY 430 TrpPheCysIleLysAsnIleAsnAspGluArg 440
DB 1022 TGGTTCTGCATCAAAATATCAACATGAAGA 1054
RESULT 10
ADN60663
ID ADN60663 standard; cDNA; 1543 BP.
XX AC ADN60663;
XX XX
DT 01-JUL-2004 (first entry)
XX DE Human secreted polynucleotide #123.
XX XX Human; secreted polynucleotide; gene; ss; autoimmune disease;
KW rheumatoid arthritis; hyperproliferative disorder; neoplasm; breast;
KW liver; cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; nervous system disorder; Alzheimer's disease;
KW bacterial infection; viral infection; fungal infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin aging; sunburn; organ transplantation; tissue regeneration;
KW chemotaxis; food additive; food preservative; fat content; vitamins;
KW mineral.
XX OS Homo sapiens.
XX XX
PN US2004038277-A1.
XX

PD 26-FEB-2004.
XX 18-JUL-2003; 2003US-00621401.
XX 07-NOV-1997; 97US-0064900P.
PR 07-NOV-1997; 97US-0064908P.
PR 07-NOV-1997; 97US-0064911P.
PR 07-NOV-1997; 97US-0064912P.
PR 07-NOV-1997; 97US-0064983P.
PR 07-NOV-1997; 97US-0064984P.
PR 07-NOV-1997; 97US-0064985P.
PR 07-NOV-1997; 97US-0064987P.
PR 07-NOV-1997; 97US-0064988P.
PR 07-NOV-1997; 97US-0066089P.
PR 07-NOV-1997; 97US-0066090P.
PR 07-NOV-1997; 97US-0066094P.
PR 07-NOV-1997; 97US-0066095P.
PR 07-NOV-1997; 97US-0066100P.
PR 04-NOV-1998; 98WO-US023435.
PR 05-MAY-1999; 99US-00305736.
PR 13-OCT-2000; 2000US-0239893P.
PR 28-MAR-2001; 2001US-00818683.
PR 12-OCT-2001; 2001US-00974879.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Feng P, Ruben SM, Ebner R, Olsen H, Ni J, Wei Y;
PI Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;
PI Endress GA, Carter KC, Birse CE;
XX WPI; 2004-203215/19.
DR P-PSDB; ADN60791.
XX New nucleic acid molecule encoding one of 125 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
PS Claim 1; SEQ ID NO 133; 531bp; English.
XX The invention relates to isolated nucleic acid molecules encoding human
CC secreted proteins. The polynucleotides and polypeptides are used in
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition. Antibodies to the polypeptides can also be used in alleviating
CC symptoms associated with the disorders and in diagnostic immunoassays
CC e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
CC Disorders which are diagnosed or treated include autoimmune diseases
CC (e.g. rheumatoid arthritis), hyperproliferative disorders (e.g. neoplasms
CC of the breast or liver), cardiovascular disorders (e.g. cardiac arrest),
CC cerebrovascular disorders (e.g. cerebral ischaemia), nervous system
CC disorders (e.g. Alzheimer's disease), infections caused by bacteria,
CC viruses and fungi and ocular disorders (e.g. corneal infection). The
CC polypeptides can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues and in chemotaxis. The polypeptides can also be
CC used as food additives or preservatives to increase or decrease storage
CC capabilities and fat content, as vitamins, as minerals and as other
CC nutritional components. This sequence represents a human secreted
CC polynucleotide of the invention.
XX
SQ Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.: 0 Length: 1543
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DB: 12 Gaps: 0

US-10-028-384-2 (1-826) x ADN60663 (1-1543)

QY 90 PheAlaValIleArgPheGluSerIleHisGluPheAspProTrpPheAsnTyrArg 109

QY 290 ValTVrileAlaTVrSerThrPheTVrileValGlyLeuileLeuSerMetGlnilePro 309
 Db 602 GTTCACATAGCATATAGCATCTTCTCATCTTGTGGTTTAAATATATATCATATGAGATACCT 561
 QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
 Db 662 TTTGTGGGATTCAGCAATCAGAACCAAGTCAACATGCGAGCTGCAGGTGTCTTTGCA 721
 QY 330 LeuLeuGlnAlaTVrAlaPheLeuGlnTVrLeuArgAspArgLeuThrLysGlnGluPhe 349
 Db 722 TTGCTGCAGCTTATGCTTTCTTCAGTATCTGAGAGACCGATTAACAAACAAGAGTTC 781
 QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
 Db 782 CAGACCCCTTTCTTTTGGGTGTATCACTAGCTGCAGGTCTGTCTTCTTATAGTGCATC 841
 QY 370 TVrLeuThrTVrThrGlyTVrileAlaProTVrSerGlyVArgPheTVrSerLeuTVrAsp 389
 Db 842 TATTGACTTATACAGTTACATTCACCATGGATGGAGGCTTTTATTCATTTGGGAT 901
 QY 390 ThrGlyTVrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
 Db 902 ACTGGGTATGCAAAATACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 961
 QY 410 ThrTVrValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
 Db 962 ACTGGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1021
 QY 430 TrpPheCysIleLysAsnIleAsnAspGluArg 440
 Db 1022 TGGTCTGCATCAAAATATCAACGATCAAGA 1054
 RESULT 9
 ADG78372
 ID ADG78372 standard; cdna; 1543 BP.
 XX AC ADG78372;
 XX DT 11-MAR-2004 (first entry)
 XX DE Human secreted protein cdna #123.
 XX KW Human; secreted protein; gene; ss; immune disorder;
 KW haematopoietic disorder; hyperproliferative disorder; infectious disease;
 KW inflammatory disorder; HIV; anaemia; thrombocytopenia; bleeding; stroke;
 KW myocardial infarction; Addison's disease; rheumatoid arthritis;
 KW dermatitis; Grave's disease; multiple sclerosis; glomerulonephritis;
 KW diabetes; graft-versus-host disease; inflammatory bowel disease; cancer;
 KW bacterial infection; viral infection; parasitic infection; osteoporosis;
 KW wound; atherosclerosis; Alzheimer's disease; Parkinson's disease; ulcer;
 KW food additive; food preservative.
 XX OS Homo sapiens.
 XX PN US2003211472-A1.
 XX PD 13-NOV-2003.
 XX PF 28-MAR-2001; 2001US-00818693.
 XX PR 05-MAY-1999; 99US-00305736.
 XX (FENG/) FENG P.
 PA (RUBE/) RUBEN S M.
 PA (ROSE/) ROSEN C A.
 PA (EBNE/) EBER R.
 PA (OLSE/) OLSEN H S.
 PA (NIJU/) NI J.
 PA (WEIY/) WEI Y.
 PA (SOFF/) SOFFET D R.
 PA (MOOR/) MOORE P A.
 PA (KYAW/) KYAW H.
 PA (LAF/) LAFLEUR D W.

(SHIY/) SHI Y.
 (JANA/) JANAT F.
 (ENDR/) ENDRESS G A.
 (CART/) CARTER K C.

Feng P, Ruben SM, Rosen CA, Ebner R, Olsen HS, Ni J, Wei Y;
 Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;
 Endress GA, Carter KC;

WPI: 2004-051575/05.
 P-PSDB: ADG78500;

New secreted nucleic acid for diagnosing, preventing or treating diseases associated with aberrant expression or activity of the polypeptide it encodes, e.g. cancer, human immunodeficiency virus, Parkinson's disease, or diabetes.

Claim 1; SEQ ID NO 133; 377pp; English.

The invention also relates to human secreted proteins and the nucleic acids encoding them. The proteins and nucleic acids are useful in diagnosing, preventing, prognosing or treating diseases or disorders associated with aberrant expression and/or activity of the secreted proteins, such as immune disorders, haematopoietic disorders, hyperproliferative disorders, infectious diseases or inflammatory disorders. In particular, the diseases or disorders are HIV, anaemia, thrombocytopenia, bleeding, stroke, myocardial infarction, Addison's disease, rheumatoid arthritis, dermatitis, Grave's disease, multiple sclerosis, glomerulonephritis, diabetes, graft-versus-host disease, inflammatory bowel disease, cancer, bacterial infections, viral infections, parasitic infections, osteoporosis, wounds, atherosclerosis, Alzheimer's disease, Parkinson's disease or ulcers. The sequences may also be used as food additives or preservatives, or for modulating mammalian mental or physical characteristics. The nucleic acids are also used in chromosome mapping, in forensic biology or as molecular weight markers. This sequence represents cDNA encoding a human secreted protein of the invention.

Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;

Alignment Scores:

Pred No.:	0	Length:	1543
Score:	351.00	Matches:	351
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	42.49%	Indels:	0
DB:	12	Gaps:	0

US-10-028-384-2 (1-826) x ADG78372 (1-1543)

Qy	90	PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTVrArg	109
Db	2	TTGCGGTATCCGCTTCGAAAGCATCATCCACGAGTTCGACCCGTGGTTTAACTATAGA	61
Qy	110	SerThrHisHisLeuAlaSerHisGlyPheTVrGluPheLeuAsnTrpPheAspGluArg	129
Db	62	TCAACACATCATCTTCATCTCATGGTTCTATGAATTTTAAATTTGTTGATGAAGA	121
Qy	130	AlaTrpTVrProLeuGlyVArgIleValGlyGlyThrValTVrProGlyLeuMetIleThr	149
Db	122	GCATGGTATCCCATAGGAAGAATAGTAGGTGTACTGTTTACCCAGGGTTGATGATACC	181
Qy	150	AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal	169
Db	182	GCTGGCCTTATTTCATTGGATTTTAAATACATTCAGAACATACTGTTCACATAAGACGTA	241
Qy	170	CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr	189
Db	242	TGTGTGTCTTCCTGCACCACTTTTAGCGCCCTTACATCATATATCTCTCTCTCTTACA	301
Qy	190	ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro	209
Db	302	AGAGACTTTGGAACCAAGGAGGAGGAGCTTTTAGCTGTCTTTTATTGCTATTGTACCA	361

RESULT 7

AAx85055
ID AAX85055 standard; DNA; 1543 BP.

XX AC AAX85055;

XX DT 30-JUL-1999 (first entry)

XX DE Human secreted protein gene No. 123.

XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
XX KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
XX KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
XX KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
XX KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
XX KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
XX KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
XX KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX OS Homo sapiens.

XX PN WO924836-A1.

XX XX 20-MAY-1999.

XX PF 04-NOV-1998; 98WO-US023435.

XX PR 07-NOV-1997; 97US-0064900P.

XX PR 07-NOV-1997; 97US-0064908P.

XX PR 07-NOV-1997; 97US-0064911P.

XX PR 07-NOV-1997; 97US-0064912P.

XX PR 07-NOV-1997; 97US-0064983P.

XX PR 07-NOV-1997; 97US-0064984P.

XX PR 07-NOV-1997; 97US-0064985P.

XX PR 07-NOV-1997; 97US-0064987P.

XX PR 07-NOV-1997; 97US-0064988P.

XX PR 17-NOV-1997; 97US-0066089P.

XX PR 17-NOV-1997; 97US-0066090P.

XX PR 17-NOV-1997; 97US-0066094P.

XX PR 17-NOV-1997; 97US-0066095P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Feng P, Carter KC, Endress GA, Rosen CA, Ruben SM, Janat F;

XX PI Ni J, Wei Y, Moore PA, Soppet DR, Kyaw H, Lafleur DW, Olsen HS;

XX PI Shi Y, Ebner R;

XX DR WPI; 1999-337740/28.

XX DR P-PSDB; AAY27689, AAY27918, AAY27919, AAY27920, AAY27921, AAY27922,

XX DR AAY27923.

XX PT New human secreted proteins and coding sequences useful for treating

XX PT disorders of the immune system and hyperproliferative disorders.

XX PS Claim 1; Page 347; 507pp; English.

XX CC This sequence represents a nucleic acid molecule which encodes a secreted

XX CC human protein. The gene number is given in the descriptor line. The gene

XX CC can be used to generate fusion proteins by linking to the gene to a human

XX CC immunoglobulin Fc portion (e.g. AAX84924) for increasing the stability of

XX CC the fused protein as compared to the human protein only. The invention

Alignment Scores:

Pred. No.: 0 Length: 1543
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DS: 2 Gaps: 0

US-10-028-384-2 (1-826) x AAX85055 (1-1543)

Qy 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArg 109
Db 2 TTGGCGGTTCATCGCTTCGAAAGACATCATCCACGAGTTCGACCGGTGTTTAACATATAGA 61
Qy 110 SerThrHisHisLeuAlaSerHisGlyPheTrpGluPheLeuAsnTrpPheAspGluArg 129
Db 62 TCAACACATCATCTTCATCTCATGGTTCTATGAATTTTAAATGGTTTGAAGAAAGA 121
Qy 130 AlaTrpTrpProLeuGlyArgIleValGlyGlyThrValTrpProGlyLeuMetIleThr 149
Db 122 GCATGGTATCCACTAGGAAGATAGTAGGTGTGTCTGTTTACCAGGGTTGATGATAACC 181
Qy 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
Db 182 GCTGGCCTTATTCATTGGATTTTAAATACATTTGAACATACTGTTCACATAAGAGACGTA 241
Qy 170 QysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
Db 242 TGTGTGTCTCTGACCAACTTTTAGCGGCCCTTACATCTATATCTATCTCTCTCTTACA 301
Qy 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValPro 209
Db 302 AGAGAACTTTGGAAACCAAGGACGAGACTTTTAGCTGTCTGTTTATTGCTATTGTACCA 361
Qy 210 GlyTrpIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
Db 362 GGCTACATATCTCGTCAGTAGCTGATCCCTTTGATATAAGAGGCATTCCTATTTTTGA 421
Qy 230 LeuGlnPheThrTrpTrpLeuTrpValIleSerValIleThrGlySerValPheTrpThr 249
Db 422 CTTTCAGTTTCACATATCTATTTATGGGTAAATCTGTAAATACTGGGTCACTTTTGGACA 481
Qy 250 MetCysCysCysLeuSerTrpPheTrpMetValSerAlaTrpGlyGlyValPheIle 269
Db 482 ATGTGCTGCTGCTATCTCTATTTCTATATGGTCTCTGCTTGGGGGGTGTATGATTTATC 541
Qy 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTrpSerIleArg 289
Db 542 ATCAATCTTATCCACTGCATGTAATTTGTTGTTGTTTACTGATGCAGAGATACAGCAAGA 601
Qy 290 ValTrpIleAlaTrpSerThrPheTrpIleValGlyLeuIleLeuSerMetGlnIlePro 309
Db 602 GTCTATACATAGCATATAGCACTTCTACATGTGGGTTTAAATATTATCAATCAGATACCT 661
Qy 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAla 329
Db 662 TTTGTGGATTCCAGCAATCAGAACACTGAAACATGGCAGCTGCAGGTGCTTTTGA 721
Qy 330 LeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeuArgAspArgLeuThrIleGluPhe 349
Db 722 TTGCTGCAAGCTTATGCTTCTTTCAGTATCTTCGAGAGACCGATTAACAAAAACAAGAGTC 781
Qy 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
Db 782 CAGACCCCTTTCTTTTGGGTGTATCCTAGCTAGCTGAGGTGCTGTGCTTCTTAGTGTATC 841
Qy 370 TyrLeuThrTrpThrTrpIleAlaProTrpSerGlyArgPheTrpSerLeuTrpAsp 389
Db 842 TATTTGACATTATACAGGTACATGTCACCATGGAGTGCAGGTTTATTTCATTGTTGGAT 901
Qy 390 ThrGlyTrpAlaLysIleHisIleProIleAlaSerValSerGluHisGlnProThr 409

QY 430 TrpPheCysIleLysAsnIleAsnAspGluArg 440
 DB 1022 TGGTTCATCATCAAAATATCAACGATGAAAGA 1054
 RESULT 6
 ABL89850
 ID ABL89850 standard; cDNA; 1209 BP.
 XX AC ABL89850;
 XX DT 24-MAY-2002 (first entry)
 XX DE Human polynucleotide SEQ ID NO 412.
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW anti-allergic; hepatotropic; antidiabetic; antinflammatory; anti-ulcer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW neurological disease; cancer; immune disorder; cardiovascular disorder;
 XX OS Homo sapiens.
 XX WO200190304-A2.
 XX 29-NOV-2001.
 XX 18-MAY-2001; 2001WO-US016450.
 XX 19-MAY-2000; 2000US-0205515P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Birse CE, Rosen CA;
 XX WPI; 2002-122018/16.
 XX P-PSDB; ABB89441.
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders.
 XX Claim 4; SEQ ID NO 412; 2081pp + Sequence Listing; English.
 XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX SQ Sequence 1209 BP; 312 A; 222 C; 235 G; 438 T; 0 U; 2 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 1209
 Score: 351.00 Matches: 351
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 42.49% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-2 (1-826) x ABL89850 (1-1209)
 QY 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrrPheAsnTyrArg 109
 DB 2 TTGGCGGTTCATCGGTTTGGAAAGCATCATCCACGAGTTCGACCCGTGGTTTAACTATAGA 61
 QY 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrrPheAspGluArg 129
 DB 62 TCAACACATCATCTTGCATCTCATGGTTCATGAATTTTAAATGGTTGATGAAAGA 121
 QY 130 AlaTrrPrrProLeuGlyArgIleValGlyThrValTrrPrrProGlyIleuMetIleThr 149
 DB 122 GCATGGTATCCACTAGGAAGAAATAGTGGTGTCTGTTCACCCAGGGTTGATGATAACC 181
 QY 150 AlaGlyLeuIleHisTrrPrrLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
 DB 182 GCTGGCCTTATTCATGGATTTAAATACATTGAACATPACTGTTTCACATAGAGACGTA 241
 QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 DB 242 TGTGTGTTCTTTCACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTCTTACA 301
 QY 190 ArgGluLeuTrrAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
 DB 302 AGAGAACTTTGGAACCAAGGAGAGAGACTTTTAGTCTGCTTGTGTTTATGCTATTGTACCA 361
 QY 210 GlyTrrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 DB 362 GGCTACATATCTCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 421
 QY 230 LeuGlnPheThrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrr 249
 DB 422 CTTCAGTTTCATACATCTATTTATGGTAAATCTGTAAAACTGGGTGAGTTTGTGGACA 481
 QY 250 MetCysCysLeuSerTrrPheTrrMetValSerAlaTrrPrrGlyTrrValPheIle 269
 DB 482 ATGTGCTGCTGCTTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 541
 QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTrrSerLysArg 289
 DB 542 ATCAATCTTATCCACTGCAATGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 601
 QY 290 ValTrrIleAlaTrrSerThrPheTrrIleValGlyLeuIleLeuSerMetGlnIlePro 309
 DB 602 GTCTACATAGCATATAGCACTTTCTACATTTGGTGGTTTAAATATATCAATGCAGATACCT 661
 QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAla 329
 DB 662 TTTGTGGATTCAGCCCAATCAGAACAAAGTGAACACATGCGAGCTGCGAGTGTCTTTGCA 721
 QY 330 LeuLeuGlnAlaTrrAlaPheLeuGlnTrrLeuArgAspArgLeuThrLysGlnGluPhe 349
 DB 722 TTGCTGCAAGCTTATGCTTTCTTTCAGTATCTGAGAGACCGATTAACAAACAGAGTTTC 781
 QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
 DB 782 CAGACCCCTTTCTTTTGGGTGATCATAGCTAGCTGCGAGTGTCTGTTCTTGTGTCATC 841
 QY 370 TrrLeuThrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrrTrr 389
 DB 842 TATTTGATTTATACAGTTTACATTTGACCATGAGTGGAGGTTTATTCATTGTGGGAT 901
 QY 390 ThrGlyTrrAlaLysIleHisIlePrrIleIleAlaSerValSerGluHisGlnProThr 409
 DB 902 ACTGGGTATGCAAAATATACATTTCCAATTTATGTCATGATGCTGCTGAGCATCAACCTACG 961
 QY 410 ThrTrrValSerPhePhePheLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
 DB 962 ACTTGGGTGCTTTCTTTCTTTGATCTACATATTTCTTGTATGATGATGATGATGATGATGAT 1021
 QY 430 TrrPheCysIleLysAsnIleAsnAspGluArg 440
 DB 1022 TGGTTCATCATCAAAATATCAACGATGAAAGA 1054

Db 2263 TGGCACTCCCGAGCTTTGACCAACACGTAATGCTGAGATTGGAAATTAAGACATTAA 2322

QY 752 sPhelyshisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgileTyrlsVa 772

Db 2323 ATTCAAGCATTTGGAGGAAGCTTTTACATCAGAGCACTGGCTTGTGAGATATATAAGT 2382

QY 772 llyshlaProAspAsnArgGluThrLeu 781

Db 2383 GAAAGCACCTGACACACAGGAGACACTA 2410

RESULT 5

AAH33264

ID AAH33264 standard; cDNA; 1209 BP.

XX

AC AAH33264;

XX

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen encoding cDNA SEQ ID NO:320.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;

KX colorectal carcinoma; ss.

OS Homo sapiens.

XX

PN WO200122920-A2.

PD 05-APR-2001.

PF 28-SEP-2000; 2000WO-US026524.

XX

PR 29-SEP-1999; 98US-0157137P.

PR 03-NOV-1999; 99US-0163280P.

XX

PA (HUMA-) HUMAN GENOME SCI INC..

XX

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX

WP1; 2001-235357/24.

DR P-PSDB; AAG73833.

XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

PT useful for preventing, diagnosing and/or treating colorectal cancers.

XX

PS Claim 1; Page 2436; 9803pp; English.

XX

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where the

CC proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene therapy

CC and vaccine production. N and P may be used in the prevention, diagnosis

CC and treatment of diseases associated with inappropriate P expression. For

CC example, N and P may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of P by expressing inactive proteins or to

CC supplement the patients own production of P. Additionally, N may be used

CC to produce the colon cancer-associated Ps, by inserting the nucleic acids

CC into a host cell and culturing the cell to express the proteins. N and P

CC can be used in the prevention, diagnosis and treatment of colorectal

CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent

CC sequences used in the exemplification of the present invention. N.B.

CC Pages 66 to 682 and page 7053 of the sequence listing were missing at

CC time of publication, meaning no sequences are present for SEQ ID NO:1027

CC to 1052, 7921 and 7922

XX

SQ Sequence 1209 BP; 312 A; 222 C; 235 G; 438 T; 0 U; 2 Other;

Alignment Scores:

Pred. No.: 0 Length: 1209

Score: 351.00 Matches: 351

Percent Similarity: 100.00% Conservativeness: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 42.49% Indels: 0

DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x AAH33264 (1-1209)

QY 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArg 109

Db 2 TTGCCGCTCATCCGCTTCGAAAGCATCATCCACGAGTTCCGCCGCTGGTTTAACTATAGA 61

QY 110 SerThrHisHisLeuAlaSerHisGlyPheTyrlGluPheLeuAsnTrpPheAspGluArg 129

Db 62 TCAACACATCATCTTGCATCTCATGGTTCTATGAATTTTAAATTTGGTTTTCATGAAAGA 121

QY 130 AlaTrpTrpProLeuGlyArgIleValGlyGlyThrValTyrlProGlyLeuMetIleThr 149

Db 122 GCATGGTATCCACTAGAGAAATAGTAGGTGGTACTGTTTACCAGGGTTGATGATAACC 181

QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169

Db 182 GCTGGCCTTATTCATTTGGATTTTAAATACATTTGAACATTAACCTGTCACATAAGAGACGTA 241

QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189

Db 242 TGTGTGTTCCTTGCACCAACTTTTAGCGGCCCTTACATCTATATCTACTTCTCTGCTTACA 301

QY 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209

Db 302 AGAGAACTTTGGAAACCAAGGAGCAGGACTTTTAGCTGCTGTTTATTTATTTGCTATTGTA 361

QY 210 GlyTrpIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229

Db 362 GGCTACATATCTCGTCAGTAGCTGATCCCTTGTATGATGAAGGCAATGCTATTTTTGA 421

QY 230 LeuGlnPheThrTyrlTyrlLeuTrpValLysSerValLysThrGlySerValPheTrpThr 249

Db 422 CTTTCAGTTTACATCTATTTATGGGTAAATCTGTAAAACTGGGTGAGTTTGTGGACA 481

QY 250 MetCysCysLeuSerTyrlPheTyrlMetValSerAlaTrpGlyGlyTyrlValPheIle 269

Db 482 ATGTGCTGCTGCTTATCTTCTATATATGGTCTCTGCTTGGGGTGGTTATGATTTATC 541

QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrlSerLysArg 289

Db 542 ATCAATCTTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601

QY 290 ValTyrlleAlaTyrlSerThrPheTyrlleValGlyLeuIleLeuSerMetGlnIlePro 309

Db 602 GTCTACATAGCATATAGCACTTCTTACATFTGGGGTTTAAATATATCAATGCAGATACCT 661

QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAla 329

Db 662 TTTTGGGATTCAGCAATCAGAACAGTGAACATGCGCAGCTGCGAGTGTCTTTGCA 721

QY 330 LeuLeuGlnAlaTyrlAlaPheLeuGlnTyrlLeuArgAspArgLeuThrIlyGlnIlyPhe 349

Db 722 TTCTGCAAGCTTATCTTCTTGTGAGTATCTGAGAGACCGGATTAACAAAACAGAGTTC 781

QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369

Db 782 CAGACCTCTTTCTTTTGGGTGTATCCTACATGCTGAGGTGCTGTGCTCTTAGTGTCTC 841

QY 370 TyrlLeuThrTyrlThrGlyTyrlleAlaProTrpSerGlyArgPheTyrlSerLeuTrpAsp 389

Db 842 TATTGACTTATACAGGTACATTTGACCATGGAGTGGCAGGTTTTTATTCATTGGGAT 901

QY 390 ThrGlyTyrlAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409

Db 902 ACTGGGTATGCAAAAATACACATTCATTTATGATCAGTGTCTGAGCATCAACCTACG 961

QY 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429

Db 962 ACTTGGGTGCTTCTTCTTCTTGTATCTTGTATGATGATGATGATGATGATGATGATGAT 1021

Pred. No.:	0	Length:	2710
Score:	365.00	Matches:	720
Percent Similarity:	98.0%	Conservative:	0
Best Local Similarity:	98.0%	Mismatches:	7
Query Match:	44.1%	Indels:	14
DB:	10	Gaps:	0

US-10-028-384-2 (1-826) x ADD94785 (1-2710)			
QY	55	ProAlaGlyLeuSerGlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPhe	74
DB	230	CCCGGGGGCTGTCGGGGGCTTGTGCGAGCGGCGGGTGGAGTCGTTGCTCTCCTTC	289
QY	75	ThrIlePheLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArg	94
DB	290	ACATCTCTCTCTGCGCTGGCTGGCGGCTTCAGCTCGCGCTCTCTCGCGCTCATCCG	349
QY	95	PheGluSerIleIleHieGluPheAspProTrpPheAsnTrpPheSerThrHisHisLeu	114
DB	350	TTTCGAGAGCATCATCCAGAGTTCGACCGGTGGTTAACTATAGATCAACACATCATCTT	409
QY	115	AlaSerHisGlyPheTy-cluPheLeuAsnTrpPheAspGluArgAlaTrpTyProLeu	134
DB	410	GCATCTCATGGATTCTATGAGTTCTAAATGGTTTGAATGAAGAGCATGGTACCCACTG	469
QY	135	GlyArgIleValGlyGlyThrValTyProGlyLeuMetIleThrAlaGlyLeuIleHis	154
DB	470	CGAAGAATAGTGGTGGCACCGCTTACCAGGGTTGATGATAACAGCTGGCTTATTTCAT	529
QY	155	TrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspValCysValPheLeuAla	174
DB	530	TGGAATTTAAATACATTGAACATAACAGTTACATGAAGAGATGGTGTATTCTCTTGA	589
QY	175	ProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsn	194
DB	590	CCAACTTTTAGCGGCTTACATCCATATCTAGTCTCTGCTAACTAGAGAACTGTGGAAC	649
QY	195	GlnGlyAlaGlyLeuAlaAlaCysPheIleAlaIleValProGlyTyIleSerArg	214
DB	650	CAAGGAGCAGGACTCTAGCTGCTTCTCATGTCATGTCATGACAGAGATGGTGTATTCTCTG	709
QY	215	SerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTy	234
DB	710	TCAGTGGCGGATCCTTTGATATATGAGGCATTTGCCATTTTGGCTTCAGTTCACATTAC	769
QY	235	TyrLeuTrpValIleSerValIleThrGlySerValPheTrpThrMetCysCysLeu	254
DB	770	TACTTATGGGTAAAGTCTGTGAAGACCGGGTCTGGTTCTGCAAAATGTGCTGCTGTG	829
QY	255	SerTyPheTyMetValSerAlaTrpGlyGlyTyValPheIleIleAsnLeuIlePro	274
DB	830	TCATATTTCTACATGGTCTCTGCGTGGGAGGTTATGTTTCATCATCAACCTCATCCCT	889
QY	275	LeuHisValPheValIleLeuLeuMetGlnArgTySerIleArgValTyIleAlaTy	294
DB	890	CTCCATGTGTTTGTGTTGCTGTGATGAGAGGTCACAGAAGAGAGTCTACATAGCATAT	949
QY	295	SerThrPhe-TyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheG	314
DB	950	AGCAC-TTGTACATTTGGGTTAAATATATCCATGCGAGATACCTTTTGGGATTTC	1008
QY	314	nProfileArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuGlnAlaTy	334
DB	1009	GCCAAATCAGAACAGGAGCAGCATGGAGCTGCAGGTGCTCTTGGCTGTGCAAGCTTA	1068
QY	334	rAlaPheLeuGlnTyLeuArgAspArgLeuThrIleGlnPheGlnPhePheP	354
DB	1069	CGCTTTTTCAGTATCTGAGAGACCGGTGACAAAACAGGAGTTCAGACCTTTTCTT	1128
QY	354	eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyIleThrTy	374
DB	1129	TTTGGTGTCTCCTAGTGTGACGGCGCTGTGTTCTTAGTGTCTATCTATCTGACATAC	1188

QY	374	rGlyTyIleAlaProTrpSerGlyArgPheTySerLeuTrpAspThrGlyTyAlaTy	394
DB	1189	AGGTTATATTCCACCATGGAGTGGCAGGTTTATTTCACTATGGATACCTGGGTATGCAA	1248
QY	394	sIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpValSerPh	414
DB	1249	AATACACATTTCCAATTTATGGCATCAGTGTCTGAACATCAGCCTACGACATGGGTGCTTT	1308
QY	414	ePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleTy	434
DB	1309	CTTCTTTGATCTACATATTTCTGTATGATCTCTCCAGCAGGCTATGGTCTGTCATCAA	1368
QY	434	sAsnIleAsnAspGluArgValPheValAlaLeuTyAlaIleSerAlaValTyPheAl	454
DB	1369	AAATATCAACATGAAGAGATTTTTCGCTCTGTATGCGATCAGTGTGTGTCTTTG	1428
QY	454	aGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAla	474
DB	1429	CGGAGTATGTTGGGGCTGATGCTGACTCTGACCCCGGCTCTGTCATGCTGTGCGCCAT	1488
QY	474	eAlaPheSerAsnValPheGluHisTyIleGlyAspMetLeuArgGluAsnProPr	494
DB	1489	CGCTTCTCCAAATGTTTTTTCAGACACTATTTGGGGATGACATGAAGAGGAAACCCACC	1548
QY	494	oValGluAspSerSerAspGluAspAspLysArgAsnGln-GlyAsnLeuTyAspLysA	514
DB	1549	TGTGGAGACAGCAGTGTATGAGGATGACAAAAGAAACCC-AGGAAACTTGTATGACAAGG	1607
QY	514	LaGlyValValArgGlyHisAla-ThrGluGlnGluTyThr-GluGluGlyLeuGlyPr	533
DB	1608	CAGGTAAAGTGAAGAGCATGT-GACAGAGCAGAGAA-ACCTGAAGAGGGCTTGGGCC	1665
QY	533	oAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHi	553
DB	1666	CAACATCAAAAGCATTTGTACCATCTGATGCTCATGCTCTGATGATGTTGCGGTCCA	1725
QY	553	sCysThrTrpValThrSerAsnAlaTySerSerProSerValValLeuAlaSerTyRas	573
DB	1726	CTGACGTGGGTCAAGAACGCTACTCCAGTCCAGGTGGTCTTCTGCTCTCTCTACAA	1785
QY	573	nHisAspGlyThrArgAsnIleLeuAspPheArgGluAlaTyPheTrpLeuArgG	593
DB	1786	TCATGATGGTACCAGGAATATATTAGATGATTTTAGAAGACGTACTTTTGGCTGAGACA	1845
QY	593	nAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyGlyTyGlnIleAlaG	613
DB	1846	AAACACGATGAACACCGCGGTCATCTCGTGGGACTACGGCTATCAGATTGCTGG	1905
QY	613	yMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLe	633
DB	1906	CATGCCCAACAGGACCACTCTGGTGGATAACAACACCTGGAAACAACAGCCACATCGCACT	1965
QY	633	uValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyLysIleMetArgThr-LeuA	653
DB	1966	GGTGGAAACCTATGCTTCCAAATGAACGGCGCTATAAATCATGAGGTCTCTTGG	2024
QY	653	spValAspTyValLeuValIlePheGlyValIleGlyTySerGlyAspAspIleA	673
DB	2025	ATGTCGATTAATGTTGTTATTTTCGAGGAGTGTGCTGCTATTCGGGGACATATCA	2084
QY	673	enLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGlu	693
DB	2085	ACAAGTTCTCTGGATGTCAGGATAGCTGAAGGGAGCATCCCAAGACATCCGGGAAG	2144
QY	693	er-AspTyPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThrLe	712
DB	2145	G-TGACTATTTTCAAC-CCAGCAGGAGAGTTCAGAGTAGACAAAGCTGGGTCTCTACTCT	2202
QY	712	uLeuAsnCysLeuMetTyIleLysMetSerTyTyArgPheGlyGluMetGlnLeuAspPh	732
DB	2203	GTTAAATGCTCTATGATATAAATGCTACTACAGATTTGGAGAAATCGACGTAGATT	2262
QY	732	eArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleTy	752

QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 Db 208 GGGCTGTGCAGCCGGCTGGGTCGGAGTCGGCTTCTCTCCCTTCCACATCTCTCTCCGGCC 267
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100
 Db 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAAAGCATCATCCAC 327
 QY 101 GluPheAspProTrpPheAsnTyrArgSerThrHisIleLeuAlaSerHisGlyPheTyr 120
 Db 328 GAGTTTCGACCCGGTGTAACTATATGATCAACATCATCTTCATCTCATGGTTCAT 397
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGlyGly 140
 Db 388 GAATTTTAAATTTGGTTTTCATGAAAGAGCATGGTATCCACTAGGAAGATAGTAGTGGT 447
 QY 141 ThrValTyrProGlyLeuMetIlePheAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
 Db 448 ACTGTTTACCCAGGCTTATATACACCGCTGGCTTATCATTTGGATTTTAAATATCATTTG 507
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 Db 508 AACATAACTGTTTACATAAGAGACGTGTGTGTCTCTTGACCAACTTTTAGCGCCTT 567
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
 Db 568 ACATCTATATCTACTTTCTCTGTTTACAAGAGACTTTTGGAAACCAAGAGGAGGACTTTTA 627
 QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
 Db 628 GCTGCTGTGTTTATTTGCTATTGTACAGGCTACATATCTCGTTCAGTAGTGGTCCCTTT 687
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleSer 240
 Db 688 GATAATGAAGGCAATGCTATTATTTGCACTTCAGTTCACATATTTATGGGTAATCT 747
 QY 241 ValIleThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetVal 260
 Db 748 GTAAAACTGGCTCAGTTTGTGGACAATGTGCTGTCTATCTATCTATATCTATATGCTC 807
 QY 261 SerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProIleHisValPheValLeu 280
 Db 808 TCTGCTTGGGGTGGTATGATTTATCATCAATCTTATTCACATGCAATGATTTGTTGTTG 867
 QY 281 LeuLeuMetGlnArgTyrSerIleArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
 Db 868 TTACTCATGCAGATACAGCAAAAGAGCTTACATAGCATATAGCACTTTCTACATTTGTG 927
 QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 Db 928 GGTTTAATATTAATCAATGCAATACCTTTTGGGATTCAGCCAAATCAGAAAGTGAA 987
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
 Db 988 CACATGGCAGCTGCGAGGTCTTTGCAATGTCGCAAGCTTATGCTTCTTCTGCAATCTG 1047
 QY 341 ArgAspArgLeuThrIleGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 Db 1048 AGAGACCGATTAACAAACAGAGTTCCAGACCTTTCTTTTGGGTGATCAGTAGCT 1107
 QY 361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrp 380
 Db 1108 GCAGGTGCTGCTTCTTCTAGTGCATCATTTTGTATCTTATACAGTTTACATGCAACATGG 1167
 QY 381 SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaIleHisIleProIleIle 400
 Db 1168 ATGGCAGGTTTATCTATTTGGGATCTGGGTATGCAAAATATACATTCATCAATATT 1227
 QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIle 420
 Db 1228 GCATCAGTGTCTGAGCATCAACTACGACTTGGGTGTCTTCTCTTTGATCTACATATT 1287

QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleIleAsnIleAspGluArg 440
 Db 1288 CTTGTATGTACCTTCCAGCAGCCCTTGGTTCATCAAAATATATCAACGATGAAGA 1347
 QY 441 ValPhe 442
 Db 1348 GTATTT 1353
 RESULT 4
 ADD94785
 ID ADD94785 standard; cDNA; 2710 BP.
 XX
 AC ADD94785;
 DT 29-JAN-2004 (first entry)
 XX
 DE Mouse SIMP cDNA sequence.
 XX
 KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA; cancer;
 KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 KW murine; gene; ss.
 XX
 OS Mus musculus.
 XX
 PN WO2003054008-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 18-DEC-2002; 2002WO-CA001967.
 XX
 PR 20-DEC-2001; 2001US-00028384.
 XX
 PA (COMP-) COMPATIGENE INC.
 XX
 PI Perreault C, McBride K;
 XX
 DR WPI; 2003-559122/52.
 XX
 PT P-FSDB; ADD94785.
 XX
 PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 or breast cancer, or for suppressing an immune response in an autoimmune
 disease.
 XX
 PS Disclosure; SEQ ID NO 3; 56pp; English.
 XX
 CC This invention relates to a novel isolated or purified human protein,
 termed source of immunodominant major histocompatibility complex (MHC) -
 associated peptide (SIMP), which is expressed ubiquitously in human
 cells, where the protein has the potential of generating several protein
 fragments binding with high affinity to a human leukocyte antigen (HLA)
 molecule. The invention may allow development of therapeutics with
 cytostatic or immunosuppressive activity or provide sequences useful for
 antisense therapy or gene therapy. The source of immunodominant MHC-
 associated peptide (SIMP) nucleic acids, proteins and fragments are
 useful for diagnosing and treating cancers, for example lung cancer,
 intestine cancer, sarcoma, prostate cancer, testicular cancer, breast
 cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 proteins are also useful for modulating an immune response. Decreasing
 lymphoid cell proliferation is useful for suppressing an immune response
 responsible for an autoimmune disease or a transplant rejection. The
 present sequence is that of the mouse SIMP cDNA which is related to the
 invention. Note: This sequence does not appear in the specification but
 was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2710 BP; 678 A; 648 C; 689 G; 695 T; 0 U; 0 Other;
 Alignment Scores:

Db 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGGTCATCCGCTTCGAAAGCATCATCCAC 327
 QY 101 GluPheAspProTyrPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyr 120
 Db 328 GAGTTCGACCGCTGTTTAACTATAGATCAACACATCATCTTCATCTCATGGTCTCAT 387
 QY 121 GluPheLeuAsnTyrPheAspGluArgAlaTyrPyrProLeuGlyArgIleValGlyGly 140
 Db 388 GAATTTTAAATGTTTGAAGAAGACATGGTATCCATAGAGAAATAGTAGGTGGT 447
 QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTyrPyrLeuAsnThrLeu 160
 Db 448 ACTGTTTACCCAGGCTTCATACACCGCTGGCTTATTCATTCGATTTTAAATACATG 507
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 Db 508 AACATAACTGTTTCAATAGAGACGTGTGTGTCTTCCTTGCCACCACTTTTAGCGGCTT 567
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuThrPheAsnGlnGlyAlaGlyLeuLeu 200
 Db 568 ACATCTATATCTTCTCTCTTACAAAGAGAACTTTTGAACCAAGGAGCAGACTTTTA 627
 QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
 Db 628 GTGCTGTTGTTTATGCTATTGTACAGGCTACATATCTCGTCAGTAGCTGGATCTTT 687
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValLysSer 240
 Db 688 GATAATGAAGCATTTGCTATTTTGCACCTTCAGTTCACATATTTATGGTAAATCT 747
 QY 241 ValLysThrGlySerValPheThrThrMetCysCysLeuSerTyrPheTyrMetVal 260
 Db 748 GTAATAACATGGCTCAGTCTTTTGGCAATGTGCTGCTTATCTTATCTTATATATGTC 807
 QY 261 SerAlaTyrGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
 Db 808 TCTGCTGGGTGGTGTATTTATATATCAATCTTATTCACATGCGATGATTTGTGTG 867
 QY 281 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
 Db 868 TACTGATGACAGATACAGAAAGAGCTTACATAGCATATACATCTTCTCATATGGT 937
 QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 Db 928 GGTTTAATATTTATCAATGACATACCTTTTGTGGATTCCAGCCAAATCAGAACAGTGAA 987
 QY 321 HisMetAlaAlaGlyValPheAlaLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
 Db 989 CACATGGCAGCTGCGAGGTCTTTTGCATGTCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1047
 QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 Db 1048 ACAGACCGATTAAACAAACAGAGTTCAGACCCCTTTTCTTTTGGGTGTATCACTAGCT 1107
 QY 361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyr 380
 Db 1108 GCAGGTGCTGCTGCTTCTTAGTGTCACTATTATGATCTATACAGTTTACATGACCATGG 1167
 QY 381 SerGlyArgPheTyrSerLeuTyrAspThrGlyTyrAlaLysIleHisIleProIleIle 400
 Db 1168 AGTGGCAGGTTTATTCATTGTTGGGATCTGGGTATGCAAAATACACATTCCAATATT 1227
 QY 401 AlaSerValSerGluHisGlnProThrThrThrValSerPhePheAspLeuHisIle 420
 Db 1228 GCATCAGTGTCTGAGATCAACCTACGACTTGGGTGTCTTTCTTTCTTGTATCTACATATT 1287
 QY 421 LeuValCysThrPheProAlaGlyLeuTyrPheCysIleLysAsnIleAsnAspGluArg 440
 Db 1288 CTTGTATGTATCTTCCAGCAGGCTTTTGGTGTCTTGCATCAAAATATCAACGATGAAGA 1347
 QY 441 ValPhe 442
 Db 1348 GTATTT 1353

RESULT 3

ADL30661

ID ADL30661 standard; cDNA; 1664 BP.

XX AC ADL30661;

XX XX

XX 20-MAY-2004 (first entry)

XX DE

XX Full length human cDNA clone seqid 2694.

XX DE

XX human; medicine; signal transduction; glycoprotein; transcription;

XX KW

XX oligo-capping method; ss; gene.

XX XX

XX OS Homo sapiens.

XX XX

XX PN EPI396543-A2.

XX XX

XX 10-MAR-2004.

XX XX

XX 07-JUL-2000; 2003EP-00025638.

XX XX

XX 08-JUL-1999; 99JP-00194486.

XX PR

XX 11-JAN-2000; 2000JP-00118774.

XX PR

XX 02-MAY-2000; 2000JP-00183865.

XX PR

XX 07-JUL-2000; 2000EP-00114089.

XX XX

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX PA

XX Ota T, Nishikawa T, Isogai T, Hayaashi K, Ishii S, Kawai Y;

XX PI

XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX XX

XX WPI; 2004-204755/20.

XX DR

XX P-PSDB; ADL30662.

XX XX

XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full

XX PT

XX length human cDNAs.

XX XX

XX Example 1; SEQ ID NO 2694; 1340pp; English.

XX PS

XX This invention relates to a novel primers useful for synthesizing full

XX CC

XX length cDNA molecules that encode human proteins. Specifically, it refers

XX CC

XX to secretory or membrane proteins that are potential therapeutic agents/

XX CC

XX target molecules in the field of medicine, and in particular genes

XX CC

XX encoding proteins that are associated with signal transduction.

XX CC

XX glycoproteins and transcription. The present invention describes a method

XX CC

XX for efficiently cloning a full length human cDNA from both the 5' and 3'

XX CC

XX ends using the oligo-capping method. This polynucleotide sequence is a

XX CC

XX full length human cDNA clone of the invention.

XX SQ

XX Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;

XX XX

Alignment Scores:

Pred. No.: 0 Length: 1664

Score: 442.00 Matches: 442

Percent Similarity: 100.00% Conservatives: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 53.51% Indels: 0

DB: 12 Gaps: 0

US-10-028-384-2 (1-826) x ADL30661 (1-1664)

QY 1

QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTyr 20

Db 28

Db 28 ATGGCGGAGCCCTCGGCCCGGAGAGACAGCAAGTCGTCCTCAACTGCTCCCGCTGG 87

QY 21

QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40

Db 88

Db 88 AGTGGCTCTATGGCCCTTGGGAAACAGCCGACCGCCACCGGGCCCGAGTGC 147

QY 41

QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60

Db 148

Db 148 CGCACAAAGCG 207

XX Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 DR P-PSDB; ADD94784.
 XX
 PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 PS Claim 6; SEQ ID NO 1; 66pp; English.

XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.

XX SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2481
 Score: 826.00 Matches: 826
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x ADD94783 (1-2481)

QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProtrp 20
 DB 1 ATGGCGAGCCCTCGCGCCCGGAGAGACAAAGTCGTCCTCAACTCGTCCCGTGG 60
 QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 DB 61 AGTGGCCTCATGGCCCTGGGNAACACCGGCACCGCCACCGGGCCCGGGCCAGTGC 120
 QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 DB 121 GGCACAAAGCG 180
 QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuSerLeuSerPheThrIleLeuPheLeuAla 80
 DB 181 GGGCTGTCGACCGCGCGCTGGTGGCAGTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
 QY 81 TrpLeuAlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100
 DB 241 TGGCTTGGCGGCTTCAGCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
 QY 101 GluPheAspProTrpPheAsnTrpArgSerThrHisHisLeuAlaSerHisGlyPheTrp 120
 DB 301 GAGTTCGACCGCGTGGTTTAACTATAGATCAACACATCATCTTGCATCTCATCGGGTCTAT 360
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGlyGly 140
 DB 361 GAATTTTAAATGGTTTCATGAAGAGCATGGTATCCATCAGGAGGAATAGTAGGTGGT 420
 QY 141 ThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160

DB 421 ACTGTTTACCAGGGTGGATGATACCGCTGCGCTTATTCATGGATTTTAAATACATTG 480
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 DB 481 AACATAACTGGTTTCACATAGACAGCATGTGTGTTCCTTGCACCAACTTTTAGCGCGCTT 540
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAlaGlyLeuLeu 200
 DB 541 ACATCTATATCTACTTCTCTTACAGAGAGACTTTGGACCAAGGAGGAGGAGCTTTTA 600
 QY 201 AlaAlaCysPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySerPhe 220
 DB 601 GCTGCTGTGTTTATTCCTATTGTACAGGCTACATATCTCGTCAGTAGTACGTCCTTT 660
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpValLysSer 240
 DB 661 GATAATGAAGGATTCCTATTTTGCATTCAGTTCACATATCTATTTAAGGTAAATCT 720
 QY 241 ValLysThrGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMetVal 260
 DB 721 GTAAAAAAGCTGGTTCAGTTCCTTTTGGACAAATGCTGCTGCTTATCTATTTCTATATG 780
 QY 261 SerAlaTrpGlyGlyValPheIleLeuLeuIleProLeuHisValPheValLeu 280
 DB 781 TCTGCTTGGGTGGTGTATGTTATTCATCAATCTTATTCACGTCGATGTTTGTGTG 840
 QY 281 LeuLeuMetGlnArgTrpSerLysArgValTrpIleAlaTrpSerThrPheTrpIleVal 300
 DB 841 TTACTGTATGTCAGATACAGCAAAAGAGTCTACATAGCATATAGCCTTTCTCATATG 900
 QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 DB 901 GGTTTATATATATCAATGACATACCTTTTGGGATTCAGCCCAATCAGAACAAAGTAA 960
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeu 340
 DB 961 CACATGGCAGCTGCGAGGTGCTTTCATTCGTCGCAAGCTTATGCTTCTTCGAGTATCTG 1020
 QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 DB 1021 AGAGCCGATTAACAAACACAGAGTTCAGACCCCTTTCTTTTGGGTGTATCACTAGCT 1080
 QY 361 AlaGlyAlaValPheLeuSerValIleTrpLeuThrTrpGlyTrpIleAlaProTrp 380
 DB 1081 CGAGTGTGCTGTTCCTTAGTGTATCTATTTGACTATACAGGTACATTCGACCAATGG 1140
 QY 381 SerGlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaLysIleHisIleProIle 400
 DB 1141 AGTGGCAGGTTTATTCATTTGGGATCTGCGGTATGCAAAAAATACACATTCCAATATT 1200
 QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIle 420
 DB 1201 GCATCAGTGTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTCTGTATCATATT 1260
 QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
 DB 1261 CTTGTATGTACTTCCCGCAGCGCTTTGGTTCCTGATCAAAAATATCAACGATGAAGA 1320
 QY 441 ValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetValArgLeu 460
 DB 1321 GTATTTGCTGCTATATGCAATCAGTGTGCTCTCTTGTGTCAGTGTGGTGGCGACTG 1380
 QY 461 MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
 DB 1381 ATGTTGACTTTTGACTCCAGTCGTGTGTATGCTGCTGCAATTCCTTTTCAATGTTTTT 1440
 QY 481 GluHisTrpLeuGlyAspMetLysArgGluAsnProProValGluAspSerSerAsp 500
 DB 1441 GAGCATTATTTGGGGATGACATGAAGGAGGAAAATCCACCTGTGGAGCAGCAGTGTAT 1500
 QY 501 GluAspLysArgAsnGlnGlyAsnLeuTrpAspLysAlaGlyValArgLysHis 520

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57; Search time 951.69 Seconds

(without alignments)

4556.130 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 826

Sequence: 1 MAEPAPESKHSKSLNSPW.....GYIKNKLVPKKGKISKKTIV 826

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+P2n.model -DEV=xih
-Q/cgn2_1/USPTO.spool/US10028384/runat_14122004_131645_6790/app_query.fasta.1.3740
-DB=N Geneseq 23Sep04 -QMT=fastcap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN 1.1.1986 @runat 14122004_131645_6790 -ICPU=3
-NO WMAP -LARGESQUEW -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database: N Geneseq 23Sep04:*

1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002as:.*
7: Geneseqn2002bs:.*
8: Geneseqn2003as:.*
9: Geneseqn2003bs:.*
10: Geneseqn2003cs:.*
11: Geneseqn2003ds:.*
12: Geneseqn2004as:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	826	100.0	2481	10	ADD94783	ADD94783 Human SIM
2	442	53.5	1664	4	AAK94164	AAK94164 Human ful
3	442	53.5	1664	12	ADL30661	ADL30661 Full leng
4	365	44.2	2710	10	ADD94785	ADD94785 Mouse SIM
5	351	42.5	1209	4	AAH33264	AAH33264 Human col
6	351	42.5	1209	6	ABL9850	ABL9850 Human pol

7	351	42.5	1543	2	AAK95055	AAK95055 Human sec
8	351	42.5	1543	8	ACD18981	ACD18981 Novel hum
9	351	42.5	1543	12	ADG78372	ADG78372 Human sec
10	351	42.5	1543	12	ADN60663	ADN60663 Human sec
11	323	39.1	2546	5	AAV44866	AAV44866 Clone CTS
12	323	39.1	2546	2	AAV98463	AAV98463 Human cDN
13	317	38.4	2510	5	AAF93772	AAF93772 Human cDN
14	297	36.0	2537	4	AAO08289	AAO08289 Human sec
15	286	34.6	2284	4	AAH18021	AAH18021 Human cDN
16	217	26.3	1114	4	AAH99794	AAH99794 Human pro
17	196	23.7	2547	4	AAO08315	AAO08315 Human sec
18	196	23.7	2660	6	ABQ54750	ABQ54750 Human ova
19	168	20.3	507	3	AAO01967	AAO01967 Human sec
20	167	20.2	787	4	AAH07526	AAH07526 Human cDN
21	166	20.1	500	6	ABV88332	ABV88332 Human col
22	145	17.6	764	5	AAF93958	AAF93958 Primer sp
23	143	17.3	433	6	ABL37128	ABL37128 Human col
24	133	16.1	558	5	AAF93612	AAF93612 Umbilical
25	119	14.4	387	4	ABA08397	ABA08397 Human sec
26	117	14.2	616	3	AAZ80248	AAZ80248 Human col
27	110	13.3	483	3	AAA43486	AAA43486 Mouse sec
28	108	13.1	485	9	ACH24071	ACH24071 Human adu
29	107	13.0	349	2	AAV89737	AAV89737 EST clone
30	94	11.4	313	8	AB220328	AB220328 Group III
31	82	9.9	250	3	AAO01440	AAO01440 Human sec
32	78	9.4	245	3	AAO17015	AAO17015 Human sec
33	78	9.4	247	2	AAK40252	AAK40252 Human sec
34	77	9.3	281	3	AAA45221	AAA45221 Human sec
35	62	7.5	388	5	AAF66328	AAF66328 Novel hum
36	54	6.5	474	5	AAK93633	AAK93633 Umbilical
37	51	6.2	393	4	ABA08547	ABA08547 Human sec
38	50	6.1	503	4	AAK91644	AAK91644 Human cDN
39	50	6.1	503	4	AAK93577	AAK93577 Human cDN
40	50	6.1	503	12	ADL30004	ADL30004 3' end of
41	50	6.1	503	12	ADL28071	ADL28071 5' end of
42	44	5.3	737	4	AAI96771	AAI96771 Human neu
43	42	5.1	2417	10	ADD94789	ADD94789 Drosophil
44	42	5.1	2699	4	ABL13247	ABL13247 Drosophil
45	42	5.1	4922	4	ABL13246	ABL13246 Drosophil

ALIGNMENTS

RESULT 1
ADD94783
ID ADD94783 standard; cDNA; 2481 BP.

XX AC ADD94783;

XX DT 29-JAN-2004 (first entry)

XX DE Human SIMP cDNA sequence.

XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;
XX KW major histocompatibility complex; human leukocyte antigen; HLA;
XX KW cytosolic; immunosuppressive; antisense therapy; gene therapy; cancer;
XX KW lung cancer; intestine cancer; sarcoma; prostate cancer;
XX KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
XX KW hematological cancer; immune response; lymphoid cell proliferation;
XX KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
XX KW gene; ss.

XX OS Homo sapiens.

XX FN WO2003054008-A2.

XX FD 03-JUL-2003.

XX PF 18-DEC-2002; 2002WO-CA001967.

XX XX 20-DEC-2001; 2001US-00028384.

XX PA (COMP-) COMPATIGENE INC.

AUTHORS **TITLE** **JOURNAL** **COMMENT**

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers
1..798
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30659262"
/tissue_type="Upper Head"
/dev_stage="embryo 9.5 - 10.5 dpc"
/lab_host="pDH10B (T1 phage resistant)"
/clone_lib="NTH-BMAP HQ0"
/note="Organ: Upper Head; Vector: pYX-Asc; Site: 1: EcoR I;
Site: 2: Not I; The library was constructed according
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CGAAGTGAAT. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

ORIGIN

Query Match 27.2%; Score 674.6; DB 7; Length 798;
Best Local Similarity 91.0%; Pred. No. 3.8e-168;
Matches 727; Conservative 0; Mismatches 71; Indels 1; Gaps 1;
QY 500 GAGAGTATGTGTCTCTTGCACCACTTTAGCGGCTTACATCTATATCTACTTTCC 559
DB 1 GAGATGTGTGTATTCCTTGCACCACTTTAGCGGCTTACATCCATATCTACGTTCC 60
QY 560 TCGTTACAGAGACTTTGGACCAAGGAGCAGACTTTAGCTGCTTTTATTGCTA 619
DB 61 TCGTAAGTACAGACTTGGACCAAGGAGCAGACTTTAGCTGCTTTTATTGCTA 120
QY 620 TTGTACCAAGGCTACATATCTCGGTACGAGTGGATCCTTTGATAATGAAGGCAATGCTA 679
DB 121 TCGTACCAAGGCTACATATCTCGGTACGAGTGGATCCTTTGATAATGAAGGCAATGCTA 180
QY 680 TTTTTCGACTTCAGTTCACACTATTATTTAGGGTAAATCTGTAAACCTGGGTCAGTTT 739
DB 181 TTTTTCGCTTCAGTTCACACTATTATTTAGGGTAAATCTGTAAACCTGGGTCAGTTT 240
QY 740 TTTGACAAATGTGCTGCTTATCTCTATTTCTATATGGTCTCTGCTTGGGTGGTTATG 799
DB 241 TCTGGACAAATGTGCTGCTTGTTCATATTTCTACATGCTCTCTGCTGGGGAGGTTATG 300
QY 800 TATTTATCATCAATCTTATTCACCTGCATGATTTGTTGTTACTGATGCAGAGATACA 859
DB 301 TGTTCATCATCAACCTCATCCCTCTCCATGTTTGTGCTGCTGATGCAGAGGTACA 360
QY 860 GCAGAAGACTTACATAGCATATAGCACTTTCTCATCTGTGGGTTTAAATATCAATGC 919
DB 361 TCAAGAGAGTCTACATAGCATATAGCACTTTCTCATCTGTGGGTTTAAATATCAATGC 420

QY 920 AGATACCTTTTGTGGGATTCAGCCCAATCAGAACAGTGAACACATGGCAGCTCCAGGTG 979
DB 421 AGATACCTTTTGTGGGATTCAGCCCAATCAGAACAGGCGACACATGGCAGCTCCAGGTG 480
QY 980 TCTTTGCATTTGCTGCAAGCTTATGCTTTCTTGCAGTATCTCGAGAGACCGATTACAAAAAC 1039
DB 481 TCTTTGCGCTGCTGCAAGCTTACGCTTTTTCAGTATCTCGAGAGACCGGTTGACAAAAC 540
QY 1040 AAGAGTTCAGACCCCTTTCTTTTGGGTGATCATCTAGCTGCAGGTGCTGTGTTCTTA 1099
DB 541 AAGAGTTCAGACCCCTTTCTTTTGGGTGATCTACTAGCTGCAGGCGCTGTGTTCTTA 600
QY 1100 GTGTCTATTTTGAATTTATACAGGTTATACAGTTTACATTTGCCACCATGGAGTGGCAGGTTTATTCAT 1159
DB 601 GTGTCTATTTTGAATTTATACAGGTTATACAGTTTATTTGCCACCATGGAGTGGCAGGTTTATTCAT 660
QY 1160 TGTGGGATTTGGGTATGCAAAATACACATTCGAATTTTGCATCAGTGTCTGAGCATC 1219
DB 661 TATGGGATTTGGGTATGCAAAATACACATTCGAATTTTGCATCAGTGTCTGAGCATC 719
QY 1220 AACTTACGACTTTGGGTGCTTTCTTTTGTGATCTACATATTTTGTATGTACCTTCCAG 1279
DB 720 AGCTTACGACATGGGTGCTTTCTTTTGTGATCTACATATTTCTGATGTACCTTCCAG 779
QY 1280 CAGGCTTTGGTTCTGCAAT 1298
DB 780 CANGGCTATGTTCTGCAAT 798

Search completed: December 15, 2004, 11:09:42
Job time : 7555.8 secs

SOURCE ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 823)
AUTHORS	NIH-MGC http://mgi.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: Dr. James Lin University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Distribution information can be found at http://genome.uiowa.edu/distribution/mousefl.html This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
FEATURES	Seq primer: pyx-5, Location/Qualifiers 1..823 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6" /db_xref="taxon:10090" /clone="IMAGE:30612197" /tissue_type="whole eye" /dev_stage="embryo 12.5,13.5,14.5 dpc" /lab_host="DH10B (TI phage resistant)" /clone_hist="NIH BMAP HD0" /note="Organ: Eye; Vector: pyx- Asc: Site 1: EcoR I; Site 2: Not I. The library was constructed according Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose Gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTAGTAGT. This library was created for the University Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the Developing Mouse Nervous System", supported by National Institute of Mental Health (NIMH)."
ORIGIN	Query Match 28.0%; Score 695.8; DB 7; Length 823; Best Local Similarity 91.1%; Pred. No. 8.5e-174; Matches 750; Conservative 0; Mismatches 72; Indels 1; Gaps 1
Qy	593 GACCTTTAGCTGCTGTTTATTCCTATGTACAGGCTACATATCTCGTCAGTAGCTG 652
Db	2 GCGTCTTAGCTCTTGCTTCATTCCTATCGTACAGGTCATATCTCGTCAGTGGCG 61
Qy	653 GATCCTCTTGATAATGAAGGCATTCCTATTTTGGCATTGATTCATCACTATTATGGG 712
Db	62 GATCCTTTGATAATGAAGGCATTCGCCATTTTGGCTTCAGTTCACTTACTTATGGG 121
Qy	713 TAAATCTGTAAATCTGGTTCAGTTTGGACATGCTGCTGCTATCTCTTCTTCT 772
Db	122 TAAATCTGTGAAGACGGGGTCTGTGTCGTGGCAATGTGCTGCTGTGCATATTTCT 181
Qy	773 ATATGGTCTCTGCTTGGGGTGGTTATGTATTTATCATCTTAATCTATTCACGTCATGTAT 832
Db	182 ACATGGTCTCTCGGTGGGAGGTTATGTGTTTCATCATCACTCATCCCTCTCCATGTGT 241
Qy	833 TTGCTGCTGCTACTGATCAGAGATACAGCAAGAGCTCTACATAGCATATTCGACTTTCT 892
Db	242 TTGCTGCTGCTGATGACAGAGGTACAGCAAGAGAGTCTACATAGCATATTCGACTTTCT 301
Qy	893 ACATTTGGGTTTAAATATTATCAATGCAGATACCTTTTGGGATTCAGCCCAATCAGAA 952
Db	302 ACATTTGGGTTTAAATATTATCAATGCAGATACCTTTTGGGATTCAGCCCAATCAGAA 361

ORIGIN

RESULT	11
L793482	
DEFINITION	UI-M_HDc ctkp-i-06-0-ui r1NH-BMAP_823 bp mRNA linear EST 10-OCT-2003 IMAGE:3061197 5' mRNA sequence.
ACCESSION	CF73482
VERSION	CF73482.1 GI:37632818
KEYWORDS	EST.

primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTATTGAAGT. This library was created for the University of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the Developing Mouse Nervous System, supported by National Institute of Mental Health (NIMH).

ORIGIN

Query Match	29.5%;	Score 731.2;	DB 7;	Length 852;
Best Local Similarity	91.0%;	Pred. No. 3.1e-183;		
Matches 775;	Conservative 0;	Mismatches 77;	Indels 0;	Gaps 0;
QY	578	GGACACAGGAGCAGGAGCTTTAGCTGCTGTTTATTTGCTATTTGACAGGCTACATAT	637	
DB	1	GGACACAGGAGCAGGAGCTTTAGCTGCTGTTTATTTGCTATTTGACAGGCTACATAT	60	
QY	638	CTCGGTCACTAGTGGATCTTTGATATGAATGAAGGCAATGCTATTTTGGCACTTCAGTTCA	697	
DB	61	CTCGGTCACTAGTGGATCTTTGATATGAATGAAGGCAATGCTATTTTGGCACTTCAGTTCA	120	
QY	698	CATACATTTTGGGTAAATCTGTAAACTGGGTCAGTTTTCGCAATGCTGCTCT	757	
DB	121	CTTACTACTTATGGTAAAGTCTGTGAAGACCGGGTCTGTGTTCTGGCAATGCTGCTCT	180	
QY	758	GCTTATCTCTATTTCTATATGCTCTCTGCTTGGGTGGTATGTTATTTATCATCAATCTTA	817	
DB	181	GCTTGTGATATTTCTACATGCTCTCTGCTGGGAGGTTATGTGTTTCATCATCACTCA	240	
QY	818	TTCCACTGCTATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	877	
DB	241	TTCCCTCTCCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	300	
QY	878	CATATAGCAGCTTTCTACATTTGGGTTTAAATATTTATCAATSCAGATACCTTTTGGGAT	937	
DB	301	CATATAGCAGCTTTCTACATTTGGGTTTAAATATTTATCAATSCAGATACCTTTTGGGAT	360	
QY	938	TCACGCCAATCAGACAGTGAACACATGCGAGCTGCGAGTGTCTTTGCTGCTGCTGCTGCT	997	
DB	361	TCACGCCAATCAGACAGTGAACACATGCGAGCTGCGAGTGTCTTTGCTGCTGCTGCTGCT	420	
QY	998	CTTATGCTTTCTTTCAGTATCTGAGAGCCGATTAACAAACCAAGAGTTCCAGACCTTTT	1057	
DB	421	CTTATGCTTTCTTTCAGTATCTGAGAGCCGATTAACAAACCAAGAGTTCCAGACCTTTT	480	
QY	1058	TCCTTTTGGGTATCTACATCTGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	1117	
DB	481	TCCTTTTGGGTATCTACATCTGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	540	
QY	1118	ATACAGGTTACATTCACCATGAGTGGGAGTTTATTCATTTGGGATCTGCTGCTGCTGCT	1177	
DB	541	ATACAGGTTACATTCACCATGAGTGGGAGTTTATTCATTTGGGATCTGCTGCTGCTGCT	600	
QY	1178	CAAAATACATTCACCATGAGTGGGAGTTTATTCATTTGGGATCTGCTGCTGCTGCTGCT	1237	
DB	601	CAAAATACATTCACCATGAGTGGGAGTTTATTCATTTGGGATCTGCTGCTGCTGCTGCT	660	
QY	1238	CTTTCTCTTTGATCTACATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	1297	
DB	661	CTTTCTCTTTGATCTACATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	720	
QY	1298	TCAAAATACATTCACCATGAGTGGGAGTTTATTCATTTGGGATCTGCTGCTGCTGCTGCT	1357	
DB	721	TCAAAATACATTCACCATGAGTGGGAGTTTATTCATTTGGGATCTGCTGCTGCTGCTGCT	780	
QY	1358	TTGCTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1417	
DB	781	TTGCTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	840	
QY	1418	CAATTGCTTTT 1429		

Db	841	GCATCGCCTTCT	852	
RESULT 9				
LOCUS	602660188F1	NCI_CGAP_Skn3	Homo sapiens	cDNA clone IMAGE:4803332 5',
DEFINITION	735 bp	mRNA	linear	EST 07-MAY-2001
ACCESSION	BC698273			
VERSION	BC698273.1	GI:13965393		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1	(bases 1 to 735)		
AUTHORS	NIH-MGC	http://mgc.ncl.nih.gov/		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: James Cleaver, M.D. cDNA Library Preparation: Life Technologies, Inc. Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10698 row: p column: 21 High quality sequence start: 2 High quality sequence stop: 734.			
FEATURES				
source	1..735			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:4803332"			
	/lab_host="DH10B (T1 phase-resistant)"			
	/clone_lib="NCI CGAP Skn3"			
	/note="Organ: skin; Vector: pCMV-SPORT6; Site: i: NotI; Site 2: SalI; Cloned unidirectionally. primer: Oligo dt. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI CGAP Library."			
ORIGIN				
Query Match	29.3%;	Score 727;	DB 4;	Length 735;
Best Local Similarity	99.3%;	Pred. No. 4e-182;		
Matches 730;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
QY	1434	TGTTTTGAGCACTATTGGGGGATGACATGAAGAGGAAATCCACCTGTGGAGCAG	1493	
DB	1	TGTTTTGAGCACTATTGGGGGATGACATGAAGAGGAAATCCACCTGTGGAGCAG	60	
QY	1494	CAGTGATGAGGATCAAAAGAAACCAAGAAATTTGTATGATAGGAGGTAAGTGAG	1553	
DB	61	CAGTGATGAGGATCAAAAGAAACCAAGAAATTTGTATGATAGGAGGTAAGTGAG	120	
QY	1554	GAAACATGCACTGAACAGGAAACCTGAAGAGGATAGCCCTATATAAAGCAT	1613	
DB	121	GAAACATGCACTGAACAGGAAACCTGAAGAGGATAGCCCTATATAAAGCAT	180	
QY	1614	TGTCACCATGTTGATGCTGATTTGATGATGATTTGCTGTCCACTGTACCTGGGTAC	1673	
DB	181	TGTCACCATGTTGATGCTGATTTGATGATGATTTGCTGTCCACTGTACCTGGGTAC	240	
QY	1674	AAGCAATGCTACTTACTGCTCAAGTGTAGTCTGCTGCTCATACATCATGATGACCCAG	1733	
DB	241	AAGCAATGCTACTTACTGCTCAAGTGTAGTCTGCTGCTCATACATCATGATGACCCAG	300	
QY	1734	GAATATCTTAGATGATTTTAGAGAGGCTTACTTTTGGCTAAGGCAAAATACAGATGAACA	1793	
DB	301	GAATATCTTAGATGATTTTAGAGAGGCTTACTTTTGGCTAAGGCAAAATACAGATGAACA	360	

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 950)
 NIH-MGC <http://mgc.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM14085 row: b column: 03
 High quality sequence stop: 680.
 Location/Qualifiers
 1..950
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6512642"
 /tissue_type="undifferentiated limb"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_134"
 /note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
 ORIGIN
 Query Match 30.9%; Score 767.6; DB 5; Length 950;
 Best Local Similarity 89.2%; Pred. No. 6.4e-193;
 Matches 838; Conservative 0; Mismatches 100; Indels 1; Gaps 1;
 643 TCAGTAGCTGGATCCTTGTGATAATGAAGGCATTCCTATTTTGACATTCAGTTCACATAC 702
 Db 1 TCAGTGGCGGATCCTTTGATATGAAGGCATTCGCCATTTTGGCTTCAGTTCACCTAC 60
 703 TATTTATGGGTAATCTGTAAATCTGGTTCAGTTCCTTTTGGCAATGCTCTCTCTTA 762
 Db 61 TACTATGGGTAATCTGTGAAGACCGGCTGTGTCTGTCTGGCAATGCTCTCTCTTG 120
 763 TCTATTCTATATGCTCTCTGCTGGGCTGTTATGTTATTCATCAATCTTATTCAC 822
 Db 121 TCATATTTCTACATGCTCTCTGGTGGGAGTTATGTTTCATCATCACTCATCCT 180
 823 CTGCATGATTTGCTGTTACTGATGAGAGATACAGCAAGAGTCTACATAGCATAT 882
 Db 181 CTTCCATGTTTGTGTTGCTGTGATGAGAGGATACAGCAAGAGTCTACATAGCATAT 240
 883 AGCATTCTTCAATCTGGGTTTATATTAATTCATTCAGATACCTTTTGGGATTCAC 942
 Db 241 AGCATTCTTCAATCTGGGTTTATATTAATTCATTCAGATACCTTTTGGGATTCAC 300
 943 CCAATCAGAACAGTGAACATGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1002
 Db 301 CCAATCAGAACAGGAGACATGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 1003 GC-TTCTTTGCAATCTGAGAGCCGATTAACAAAAAAGAGTTCAGACCCCTTTCTT 1061
 Db 361 GCTTTTTCAGATATCTGAGACCGGTTGACAAAAAGAGTTCAGACCCCTTTCTT 420
 1062 TTGGGCTGATCCTAGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1121
 Db 421 TTGGGCTGCTCCTAGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 1122 AGTTACATTCACCATGAGTGGCAGGCTTTTATTCATTTGGGATCTCGGTATGCAAA 1181
 Db 481 AGTTATATTCACCATGAGTGGCAGGCTTTTATTCATTTGGGATCTCGGTATGCAAA 540

QY 1182 AATACACATTCCAATTAATTCATCAGTCTGTCAGCATCAACCTACGACTTGGGTGCTTT 1241
 Db 541 AATACACATTCCAATTAATTCATCAGTCTGTCAGCATCAACCTACGACTTGGGTGCTTT 600
 QY 1242 CTTCTTTGATCTACATATCTTGTATGTATGATCTTCCAGCAGGCTTTGGTTCGATCAA 1301
 Db 501 CTTCTTTGATCTACATATCTTGTATGTATGATCTTCCAGCAGGCTTTGGTTCGATCAA 660
 QY 1302 AATATCAACGATGAAGAGTATTTTGTCTTATATATCAATCAGTGTCTTACTTTGC 1361
 Db 661 AATATCAACGATGAAGAGTATTTTGTCTTATATATCAATCAGTGTCTTACTTTGC 720
 QY 1362 TGAGGTGATGTTGCGACTGATCTGACCTTGTGACTCCAGTCTGTTGATGCTGTCAAT 1421
 Db 721 CNGAGTGTGTTGCGGCTGATCTGACTCTGACCCCGTCTGTCATGCTGTGCGGCT 780
 QY 1422 TGCTTTTCAATGTTTTTGGACATCTATTTGGGGGATGACATGAAAGGGAATTCAC 1481
 Db 781 CGCTTTCTCCAATGTTTTTGGACATCTATTTGGGGGATGACATGAAAGGGAATTCAC 840
 QY 1482 TGTGGAGGACAGCAGTGTGATGAGATGACAAAAGAAACCAAGAAATTTGATGATAAGGC 1541
 Db 841 TGTGGAGGACAGCAGTGTGATGAGATGACAAAAGAAACCAAGAAATTTGATGATAAGGC 900
 QY 1542 AGTTAAGTGAAGAAACATGCACTGTAACAGAGGAAAGAAAC 1580
 Db 901 AGTTAAGTGAAGAAACATGTAACAGAGGAAAGAAAC 939
 RESULT 8
 CN458732 852 bp mRNA linear EST 21-APR-2004
 LOCUS UI-M-HB0-cov-1-02-0-UI-r1 NIH_BMAP_HB0 Mus musculus CDNA clone
 DEFINITION IMAGE:30649513 5', mRNA sequence.
 ACCESSION CN458732
 VERSION CN458732.1 GI:46464458
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 852)
 NIH-MGC <http://mgc.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
 Seq primer: pyx-5.
 Location/Qualifiers
 1..852
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30649513"
 /tissue_type="whole eye"
 /dev_stage="embryo 12.5, 13.5, 14.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_HB0"
 /notes="Organ: Eye; Vector: pyx- Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT

FEATURES
source

QY 2111 TAGACAAAGCAGGATCCCTACTTTGTTGTAATGCTTATGTTATATAAATGTCATACATA 2170
 |||||
 Db 481 TAGACAAAGCAGGATCCCTACTTTGTTGTAATGCTTATGTTATATAAATGTCATACATA 540
 |||||
 QY 2171 GATTGGGAAATGAGCTGCTGATTTTGGTACACCCCGGCTTTTGACCGAACAACGTAATG 2230
 |||||
 Db 541 GATTGGGAAATGAGCTGCTGATTTTGGTACACCCCGGCTTTTGACCGAACAACGTAATG 600
 |||||
 QY 2231 CTGAGATTGGAAATGAGGACATTTAAATCAAAATTTGGGAAAGGCTTTTACATCAGAAC 2290
 |||||
 Db 601 CTGAGATTGGAAATGAGGACATTTAAATCAAAATTTGGGAAAGGCTTTTACATCAGAAC 660
 |||||
 QY 2291 ACTGCTTGTAGGATATATAAAGTAAAGCAGCTGATTAACAGGAGGAGACATTAGATCACA 2350
 |||||
 Db 661 ACTGCTTGTAGGATATATAAAGTAAAGCAGCTGATTAACAGGAGGAGACATTAGATCACA 720
 |||||
 QY 2351 AACCTCGAGTACCAACATTTTCCAAACAGAAAGTATTTGTCAAAGAGACTACCAAAA 2410
 |||||
 Db 721 AACCTCGAGTACCAACATTTTCCAAACAGAAAGTATTTGTCAAAGAGACTACCAAAA 780
 |||||
 QY 2411 GGAAGCGTGGTACATTAATAAATTAAGCTGGTTTAAAGAAAGGCAAGAAATATCTAAGA 2470
 |||||
 Db 781 GGAAGCGTGGTACATTAATAAATTAAGCTGGTTTAAAGAAAGGCAAGAAATATCTAAGA 840
 |||||
 QY 2471 AGACTGTTTAA 2481
 |||||
 Db 841 AGACTGTTTAA 851
 |||||

RESULT 6

LOCUS BX363211 1088 bp mRNA linear EST 08-APR-2004
 DEFINITION BX363211 Homo sapiens HELA CELLS cot 25-NORMALIZED Homo sapiens
 CDNA clone CS0DK010Y001 5-PRIME, mRNA sequence.

ACCESSION BX363211

VERSION BX363211.2 GI:46291519

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1088)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On May 5, 2003 this sequence version replaced gi:30372696.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

2459.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?c=CS0DK010AH01QP14c=2459.f.

Location/Qualifiers

1..1088

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DK010Y001"

/cell_type="HELA CELLS"

/cell_line="HELA"

/clone_lib="Homo sapiens HELA CELLS cot 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and EcoR V into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES

source

ORIGIN

Query Match 32.6%; Score 809.8; DB 5; Length 1088;
 Best Local Similarity 96.2%; Pred. No. 3.7e-204;
 Matches 821; Conservative 24; Mismatches 6; Indels 2; Gaps 2;
 QY 1631 TGATGCTATTGATGATTTGCTGCTCCACTGTACCTGGGTCCACAGCAATGCCCTACTCTA 1690
 |||||
 Db 1 TGATGCTATTGATGATTTGCTGCTCCACTGTACCTGGGTCCACAGCAATGCCCTACTCTA 60
 |||||
 QY 1691 GTCCTAAGTGTAGTCTGGCTCATACATCATCATGATGCGCAGCAGGAATATCTTAGATGATT 1750
 |||||
 Db 61 GTCCTAAGTGTAGTCTGGCTCATACATCATCATGATGCGCAGCAGGAATATCTTAGATGATT 120
 |||||
 QY 1751 TTAGAGAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGGAATGCAAGTATGTCCTT 1810
 |||||
 Db 121 TTAGAGAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGGAATGCAAGTATGTCCTT 180
 |||||
 QY 1811 GGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTATAGAACTACCTTGGTGGATAATA 1870
 |||||
 Db 181 GGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTATAGAACTACCTTGGTGGATAATA 240
 |||||
 QY 1871 ACACCTTGGAAATACAGCCACATAGCAGCTGGTGGGAAAAGCTATCTCTTAATGAAACAG 1930
 |||||
 Db 241 ACACCTTGGAAATACAGCCACATAGCAGCTGGTGGGAAAAGCTATCTCTTAATGAAACAG 300
 |||||
 QY 1931 CAGCCTTAAATCATGAGGACTCTAGATGTAGATATGTTTGGTTATTTTGGAGGGG 1990
 |||||
 Db 301 CAGCCTTAAATCATGAGGACTCTAGATGTAGATATGTTTGGTTATTTTGGAGGGG 360
 |||||
 QY 1991 TTATTGGCTATTCTGGTGATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAG 2050
 |||||
 Db 361 TTATTGGCTATTCTGGTGATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAG 420
 |||||
 QY 2051 GAGAAATCCCAAGACATTTGGGAAAGTGAATTTTACCCACAGGAGGAAATTCGGTG 2110
 |||||
 Db 421 GAGAAATCCCAAGACATTTGGGAAAGTGAATTTTACCCACAGGAGGAAATTCGGTG 480
 |||||
 QY 2111 TAGACAAAGCAGGATCCCTACTTTGTTGTAATGCTTATGTTATATAAATGTCATACATA 2170
 |||||
 Db 481 TAGACAAAGCAGGATCCCTACTTTGTTGTAATGCTTATGTTATATAAATGTCATACATA 540
 |||||
 QY 2171 GATTGGGAAATGAGCTGCTGATTTTGGTACACCCCGGCTTTTGACCGAACAACGTAATG 2230
 |||||
 Db 541 GATTGGGAAATGAGCTGCTGATTTTGGTACACCCCGGCTTTTGACCGAACAACGTAATG 600
 |||||
 QY 2231 CTGAGATTGGAAATGAGGACATTTAAATCAAAATTTGGGAAAGGCTTTTACATCAGAA 2289
 |||||
 Db 601 CTGAGATTGGAAATGAGGACATTTAAATCAAAATTTGGGAAAGGCTTTTACATCAGAA 660
 |||||
 QY 2290 CACTGGCTTTGATGATATATAAAGTAAAGCAGCTG-ATACAGGAGGAGACATTAGATCA 2348
 |||||
 Db 661 CACTGGCTTTGATGATATATAAAGTAAAGCAGCTG-ATACAGGAGGAGACATTAGATCA 720
 |||||
 QY 2349 CAAACCTCGAGTACCAACATTTTCCAAACAGAAAGTATTTGTCAGAAAGACTACCAA 2408
 |||||
 Db 721 CAAACCTCGAGTACCAACATTTTCCAAACAGAAAGTATTTGTCAGAAAGACTACCAA 780
 |||||
 QY 2409 AAGGAAGCTGCTGCTACATTTAAATTAAGCTGGTTTAAAGAAAGGCAAGAAATATCTAA 2468
 |||||
 Db 781 AAGGAAGCTGCTGCTACATTTAAATTAAGCTGGTTTAAAGAAAGGCAAGAAATATCTAA 840
 |||||
 QY 2469 GAAGACTGTTTAA 2481
 |||||
 Db 841 GAAGACTGTTTAA 853
 |||||

RESULT 7

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

BUS15954 950 bp mRNA linear EST 12-SEP-2002
 AGENCOURT 10120433 NIH_MGC_134 Mus musculus cDNA clone
 IMAGE:6512642 5', mRNA sequence.
 BUS15954
 BUS15954.1 GI:22823480
 EST.

/note="Vector: PCMV-SPORT6"

ORIGIN

Query Match 66.1%; Score 1640.6; DB 3; Length 4513;
Best Local Similarity 82.3%; Pred. No. 0;
Matches 1911; Conservative 0; Mismatches 404; Indels 8; Gaps 2;

```
QY 159 GGCCTCGAAGACATCATCCAGAGTTCGACCCGCTGTCGACCGGCTGGGTGGGAGTCGCTCTC 218
DB 414 GGGCCATGGAGACTGTGGGCGGTGTGTGCGACCGCGGGCTGGCAGTCTCTGCTCTC 473
QY 219 CTTACACATCTCTCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 278
DB 474 CTTACATCTCTCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 533
QY 279 CGCTTCGAAACATCATCCAGAGTTCGACCCGCTGTTAACTATATAGATCAACATCA 338
DB 534 CGCTTCGAGACATCATCCAGAGTTCGACCCGCTGTTAACTATATAGATCAACATCA 593
QY 339 TCTTGCACTCATGGGTCTATGAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 398
DB 594 CTTAGCATCCCATGCTTTTATGAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 653
QY 399 ACTAGGAAGATAGTAGGTGCTGTTTACCCAGGCTGATGATACCCGCTGGCTGCTAT 458
DB 654 ACTGGGAGATAGTAGGTGGAACAGTATACCTGGATTAATGTTACAGCTGGCTCAT 713
QY 459 TCATTGGATTTTAAATACATTCGACATACCTGTTACATGAAGAGCTATGTTGTTCT 518
DB 714 CCATTGGATTTTAAATACATTCGACATACCTGTTACATGAAGAGCTATGTTGTTCT 773
QY 519 TGCACCAACTTTAGCGGCTTACATCTATCTACTTCTGCTTACAGAGAACTTTG 578
DB 774 AGCACCAGTATTTAGTGGCTTACATCAATCTCCACTTCTGCTACTCGAAGCTGTG 833
QY 579 GAACCAAGAGCAGACATTTAGCTGCTGTTTATGCTATGCTACCGGCTACATATC 638
DB 834 GAACCAAGAGCAGACATTTAGCTGCTGTTTATGCTATGCTACCGGCTACATATC 693
QY 639 TCGGTGAGTGGTCTGTTGATTAATGAAGCAATTCCTATTTTGCACTTCAGTTAC 698
DB 894 CAGATCAGTGGTCAATTTGCAATGAAGCAATTCCTATTTTGCACTGGAGTTAC 953
QY 699 ATACTATTTATGGTAAATCTGTAATAAATCTGGGTGAGTTTGGGCAATGCTGCTG 758
DB 954 ATATTATTTGCGTAAATCTGTAATAAATCTGGGTGAGTTTGGGCAATGCTGCTG 1013
QY 759 CTTATCCTATTTCTATGCTCTGCTGGGTGGTATGATTTATATCATCAATCTTAT 818
DB 1014 TCTGCTTATTTTATATGCTTCCGCTGGGCTGTTGTTGTTATCTATTAATCTTAT 1073
QY 819 TCCACTGCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 878
DB 1074 TCCACTGCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1133
QY 879 ATATAGCACTTCTCATGTTGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 938
DB 1134 TTACAGCACTTTTACATTTGGGTTTAACTACTAATGCAAGATTCCTTTGCGGATTT 1193
QY 939 CCAGCCAAATCAGAAAGTGAACATGCGAGCTGCAAGTGTCTTTGCAATGCTGCAAGC 998
DB 1194 TCAGCCAAATCAGAAAGTGAACATGCGAGCTGCAAGTGTCTTTGCGGATTTGCGAAGC 1253
QY 999 TTATGCTTTCTTGGAGTATCTGAGAGCCGATTAACAAACAGAGTTCCAGACCTTTT 1058
DB 1254 CTATGCTTTCTTGGAGTATCTGAGAGCCGATTAACAAACAGAGTTTCAGACCTTTT 1313
QY 1059 CTTTGTGGGTGATCACTAGCTGCAAGTGTGCTGTTCTTATGCTCATCTATTGACTTA 1118
DB 1314 TTTCTCGTGTATCGTTGGCAGCTGGAGCTGTTCTCAGTGTCTATTTGACTTA 1373
QY 1119 TACAGTTCATGTCACCATGAGTGGCAGCTTTTATCTATTTGCGGATCTGGGTATGC 1178
```

```
DB 1374 TACAGGCTACATTTGCTCCATGGAGTGGAGATTTTACTCATTGCGACACAGGATGC 1433
QY 1179 AAAAAATACATTTCCAAATTTATTTGATTCAGTGTCTGAGCATCAACTAGCATTTGGGTGC 1238
DB 1434 GAAATTCATATCCCATCATTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1493
QY 1239 TTTCTTTCTTTGATCTACATATTTCTTGTATGTATGTTACCTTCCAGCAGGCTTTGGTTCGAT 1298
DB 1494 CTTCTTTTGTGATCTGATATATTAGTAGTATGATATTTCTCTCAGGACTTTGGTTCGAT 1553
QY 1299 CAAAAATATCAAGATGAAGAGTATTTGTTGCTCTATATGCAATCAGTCTGCTACTT 1358
DB 1554 AAAAAATATCAAGATGAAGAGTATTTGTTGCTCTGATGCAATTAGTCCGTTACTT 1613
QY 1359 TGTGAGTGTGATGCTGCGACTGATGTTGACTTTGACTCCAGTCTGCTGTATGCTGTCTGC 1418
DB 1614 TGTGCGGTAAATGTTGTTGTTGATGCTGACCTTGACTCTCTGTGTTGCTGCTGCGC 1673
QY 1419 AATTGCTTTTCAAAATTTTGTGCACTATTTGGGGATGACATGAAAGGAAATCC 1478
DB 1674 CATGCTCTTCTCCAAATGTTTGTGAGCATTTCTGGGTGATGATGAAAGCGGAAACCC 1733
QY 1479 ACTGTGGAGGACAGCAGTGTGATGAGGATGACAAAAAGAAACCAAGGAAATTTGTATGATA 1538
DB 1734 CCCAGTGAAGACAGCAGTGAAGAG-----AGCGAAATCTGGTACACTCTATGATA 1787
QY 1539 GGCAGGTAAAGTGAAGACATGCACTGAACAGAGAAABAACTGAAGAGGATTTAGGCC 1598
DB 1788 GGTGCAAGTGAAGACATGTCCTGAGCAGAGAAAGCAGAGGAGTCTTTGGCC 1847
QY 1599 TAATAAAAAAGCAATGTCACTATGTTGATGCTGATGTTGATGATGTTGTTGCTGCTCA 1658
DB 1848 TAATAAAGAGTATAGTCACTATGTTGATGCTGATGCTGATGATGATGATGATGATGATG 1907
QY 1659 CTGTACTGCTGCTCACAAGCAATGCTCTACTCTAGTCCAAAGTGTAGTCTGGCTCATACAA 1718
DB 1908 CTGTACTGCTGCTCACAAGCAATGCTCTACTCTAGTCCAAAGTGTAGTCTGGCTCATACAA 1967
QY 1719 TCAATGAGGACAGCAATGATCTTATGATGTTTATGAGAGCTTACTTTTGGTGAAGCA 1778
DB 1968 TCAGATGGAACAGCAATGATCTTGGATGATCTTGGATGATCTTGGATGATCTTGGATGAT 2027
QY 1779 AATACAGATGAACATGCTGAGTAAATGCTTGGTGGGATGATGCTATCAGATGCTG 1838
DB 2028 GAATACAGATGAACATGCTGAGTAAATGCTTGGTGGGATGATGCTATCAGATGCTG 2087
QY 1839 AATGCTTATGAGCAATGCTGAGTAAATGCTTGGTGGGATGATGCTATCAGATGCTG 1898
DB 2088 AATGCTTATGAGCAATGCTTGGATGATCTTGGATGATCTTGGATGATCTTGGATGAT 2147
QY 1899 GGTGGGAAAAAGCTATGCTTCTTAAATGAAACAGCAGCTTAAATATCAAGAGCTCTAGA 1958
DB 2148 GGTGGGAAAAAGCTATGCTTCTTAAATGAAACAGCAGCTTAAATATCAAGAGCTCTAGA 2207
QY 1959 TGTAGATTTGTTTGGTATTTTGGAGGGTATTTGGCTATTTCTGCTGATGATATCAA 2018
DB 2208 TGTAGATTTGTTTGGTATTTTGGAGGGTATTTGGCTATTTCTGCTGATGATATCAA 2267
QY 2019 CAAATTTCTCTGATGTTTGGATGCTCAAGAGGAGCAATCCCAAGAGCAATTCGGGAAAG 2078
DB 2268 CAAATTTCTCTGATGTTTGGATGCTCAAGAGGAGCAATCCCAAGAGCAATTCAGGAAAG 2327
QY 2079 TGTATTTTACCCACAGGAGATTCGCTGAGCAAGAGGATCCCTACTTTGTT 2138
DB 2328 TGTACTTTTACCCACAGGAGATTCGCTGAGCAAGAGGATTCCTACTTTGTT 2385
QY 2139 GAATTTGCTTTATGATAAAATGCTACTATACAGATTTGGAGAAATGAGCTGGATTTTCG 2198
DB 2386 CAATTTGCTTTATGATAAAATGCTACTATACAGATTTGGAGAAATGAGCTGGATTTTCG 2445
QY 2199 TACACCCCGAGTTTTCAGGACAGCTGATGCTGAGTATGAGATGAGGATTAATTT 2258
DB 2446 AACACACCTGGATTTTGCACCGGACGTAATCTGAAATTTGGGAAATAGGATTTTAAAT 2505
```


Db 1046 GAGAGACGGTTGACAAACAGGAGTTCACAGACCCCTTTCTTTTGGGTGTCCTCACTAGC 1105
Qy 1080 TGCAGGTGTGCTGCTTCTTAGTGTCACTATTTTCACTTATACAGGTGTACATTCACCATG 1139
Db 1106 TGCAGCGGTGTGCTTCTTAGTGTCACTATTTTCACTTATACAGGTGTATATTCACCATG 1165
Qy 1140 GAGTGGCAGTTTATTCATTTGTGGGATACCTGGGTATGCAAAATACACATTCCTCAATAT 1199
Db 1166 GAGTGGCAGTTTATTCATTTGTGGGATACCTGGGTATGCAAAATACACATTCCTCAATAT 1225
Qy 1200 TGCATCAGTGTCTGAGCATCACTTACGACTTGGGTCTTCTTCTTGTGATCTACATAT 1259
Db 1226 TGCATCAGTGTCTGAGCATCACTTACGACTTGGGTCTTCTTCTTGTGATCTACATAT 1285
Qy 1260 TCTTGTATGCTCTTCCAGCAGGCTTTGCTTCTGATCAAAATATCAACGATGAAAG 1319
Db 1286 TCTTGTATGCTCTTCCAGCAGGCTTTGCTTCTGATCAAAATATCAACGATGAAAG 1345
Qy 1320 AGTATTTGTGCTCTATATGCAATCAGTGTCTCTTCTTCTGCTGAGTGTGCTGACT 1379
Db 1346 AGTATTTGTGCTCTATATGCAATCAGTGTCTCTTCTTCTGCTGAGTGTGCTGACT 1405
Qy 1380 GATTTGTGCTCTTCACTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1439
Db 1406 GATTTGTGCTCTTCACTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1465
Qy 1440 TGAGCAGTATTTGGGGGATGATCAATGAGGAGGAAATCCACCTGTGGAGCAGCAGTGA 1499
Db 1466 TGAGCAGTATTTGGGGGATGATCAATGAGGAGGAAATCCACCTGTGGAGCAGCAGTGA 1525
Qy 1500 TGAGGATGCAAAAGAAACCAAGGAAATTTGATGATGAGGAGGAAATTTGAGGAAAC 1559
Db 1526 TGAGGATGCAAAAGAAACCAAGGAAATTTGATGATGAGGAGGAAATTTGAGGAAAC 1585
Qy 1560 TCCACTGACAGGAGGAAATCTGAGGAGGAAATTTGAGGAGGAAATTTGAGGAGGAAAT 1619
Db 1586 TGTGACAGCAGCAGGAGGAAATCTGAGGAGGAAATTTGAGGAGGAAATTTGAGGAGGAAAT 1645
Qy 1620 CATGTTGATGCTGCTATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1679
Db 1646 CATGTTGATGCTGCTATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1705
Qy 1680 TGCCTACTTCTGCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1739
Db 1706 CCGCTACTTCTGCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1765
Qy 1740 CTTAGATGATTTAGAGAGGCTTACTTTTGGCTTAAAGCAAAATACAGATGAAATGACG 1799
Db 1766 ATTAGATGATTTAGAGAGGCTTACTTTTGGCTGAGACAAACACGAGTGAACACGCGCG 1825
Qy 1800 AGTATGCTCTTGGTGGGATTTGGCTATCAGATGCTGGAATGGCTAATAGAACTACGTT 1859
Db 1826 GGTATGCTCTTGGTGGGATTTGGCTATCAGATGCTGGAATGGCTAATAGAACTACGTT 1885
Qy 1860 GGTGATATTAACCTGGAATTAACAGCCATAGCATGCTGGTGGGAAAGCTATGCTTC 1919
Db 1886 GGTGATATTAACCTGGAATTAACAGCCATAGCATGCTGGTGGGAAAGCTATGCTTC 1945
Qy 1920 TAATGAAACAGCAGGCTTAAATCATGAGGACTCTAGATGATGATGATGATGATGATGAT 1979
Db 1946 CAATGAAACAGGCTTAAATCATGAGGACTCTAGATGATGATGATGATGATGATGAT 2005
Qy 1980 TTTTGGAGGGTATGCTTATTTCTGGGATGATATCAACAAATTTCTCTGATGATGAT 2039
Db 2006 TTTTGGAGGGTATGCTTATTTCTGGGATGATATCAACAAATTTCTCTGATGATGAT 2065
Qy 2040 GATAGCTGAAGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGG 2099
Db 2066 GATAGCTGAAGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGG 2125
Qy 2100 AGAATTCGCTGAGCAAGAGGATCCCTTCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 2159
Db 2126 AGAATTCGCTGAGCAAGAGGATCCCTTCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 2185

Qy 2160 GTCTACTACAGATTGGAGAAATGCAGTGGATTTTCGTACACCCCGTGTGTGACCG 2219
Db 2186 GTCTACTACAGATTGGAGAAATGCAGTGGATTTTCGTACACCCCGTGTGTGACCG 2245
Qy 2220 AACACSTAAATGCTGAGATTGGAAATTAAGACATTAATTAATTAATTAATTAATTAAT 2279
Db 2246 AACACSTAAATGCTGAGATTGGAAATTAAGACATTAATTAATTAATTAATTAATTAAT 2305
Qy 2280 TACATCAGAACACTGGCTTGTAGGATATATAAGTAAAGCACTGATTAACAGGGAGAC 2339
Db 2306 TACATCAGAACACTGGCTTGTAGGATATATAAGTAAAGCACTGATTAACAGGGAGAC 2365
Qy 2340 ATTAGATCAAAACCTCGAGTCACCAATTTTCCAAAACAGAAATTTTGTCAAGAA 2399
Db 2366 ACTAGTCAAAACCTCGAGTCACCAATTTTCCAAAACAGAAATTTTGTCAAGAA 2425
Qy 2400 GACTACCAAAAGAGAGCGGTGCTACATTAATAAATAGCTGTTTAAAGAAAGCAAGAA 2459
Db 2426 GACTACCAAAAGAGAGCGGTGCTACATTAATAAATAGCTGTTTAAAGAAAGCAAGAA 2485
Qy 2460 AATATCTAAAGAGACTGTTTAA 2481
Db 2486 GACCTCTAAAGAGACTGTTTAA 2507

RESULT 3
BC044321

LOCUS BC044321 4513 bp mRNA linear HTC 23-JAN-2003
DEFINITION Xenopus laevis, clone IMAGE:4684177, mRNA.
ACCESSION BC044321
VERSION BC044321.1 GI:27882209
KEYWORDS HTC.

SOURCE

ORGANISM Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Xenopus.

REFERENCE

AUTHORS Klein, S. and Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (17-JAN-2003) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA

REMARK

COMMENT Contact: XGC help desk
Email: cgapps-remail.nih.gov
Tissue Procurement: Dr. Igor Dawid
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: anadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC plate: 94 Row: i Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein
This clone has the following problem: frame shifted.

FEATURES

source
1. 4513
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:4684177"
/tissue_type="Embryo, stage 31/32, Xenopus"
/clone_lib="NCHD_XGC_Emb4"
/lab_host="DH10B"

AUTHORS
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL REFERENCE
 Nature 420, 563-573 (2002)

AUTHORS
 6 (bases 1 to 2709)
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE
 Direct Submission

JOURNAL
 Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsr.riken.jp, URL: http://genome.scr.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]
 On Dec 10, 2002 this sequence version replaced gi:12858635.
 Please visit our web site (<http://genome.scr.riken.jp/>) for further details.

COMMENT
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGCGGCGAAGCGAGTGTGTTTCTTTTCTT-3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGAGATCAGAGCTCAATTAATTAACCCCGCCCGC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES
 source
 1..2709
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM_DB:1300006C19"
 /db_xref="taxon:10090"
 /clone="1300006C19"
 /sex="male"
 /tissue_type="liver"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 76..2547
 /note="unnamed protein product; hypothetical oligonucleotyl transferase (OTase) Srt3 subunit containing protein (InterPro|IPR003674, evidence: InterPro)
 putative"
 /codon_start=1
 /protein_id="BAB31390.1"
 /db_xref="GI:12858636"
 /translation="MASPSAPESKHKSLNSPWSGLMAGNSRHCHGPGTQASRA
 AAPKPGPAGLSGLSPAGMQSLFTFLFLAWLAGFSRLFAVIRPESIHEDFPW
 FNYSTHLSHSGYEFLEPNFDEAWPLGRVGVVFLMLITAGLIHLWILNLT
 VHIRDVCVFAPTSGLTSISTELLRELWNOGALLAACFIAIPVGIYSRSVSGFD
 NEGAIIFALQFTYLYWTSKYKTSVFVWCCCLSYFVMSAWGVVFINLPLHFV
 LLLMQRYSKYVAYSTLYIVGLISLQIPFVGPQIRTSSEHMAAGFALLQYAFI
 QYLRDRITKQBFQIFLFLGLAAGFVLSVLYITVTVGIAFWSGRFSLMDTGAKI
 HIPITIASVSHOPTWVSFFDFHLVCTFPAGLWFCIKINIDERVFLVALYISAVF
 AGVMVRLMLITPVVCMLSAIPSNVPEHVLGDMKRENPVEDSDSDKPNGLY
 DKAKVAKHTEQKREGLEGPNIKSIIVTMLMLLMFAVCTWTSNAYSPPSVL
 ASYHNDGTNLDDFRAIFWLNQNDDEHARVMSWDYGIQIAGMANRTLLVDNNWN
 NSHIALVGVKAVSSNETAAYKIMRSLDWDYVYLVIFGVGVIYSGDDINKFLWMVRIAGE

CDS
 /note="unnamed protein product; hypothetical oligonucleotyl transferase (OTase) Srt3 subunit containing protein (InterPro|IPR003674, evidence: InterPro)
 putative"
 /codon_start=1
 /protein_id="BAB31390.1"
 /db_xref="GI:12858636"
 /translation="MASPSAPESKHKSLNSPWSGLMAGNSRHCHGPGTQASRA
 AAPKPGPAGLSGLSPAGMQSLFTFLFLAWLAGFSRLFAVIRPESIHEDFPW
 FNYSTHLSHSGYEFLEPNFDEAWPLGRVGVVFLMLITAGLIHLWILNLT
 VHIRDVCVFAPTSGLTSISTELLRELWNOGALLAACFIAIPVGIYSRSVSGFD
 NEGAIIFALQFTYLYWTSKYKTSVFVWCCCLSYFVMSAWGVVFINLPLHFV
 LLLMQRYSKYVAYSTLYIVGLISLQIPFVGPQIRTSSEHMAAGFALLQYAFI
 QYLRDRITKQBFQIFLFLGLAAGFVLSVLYITVTVGIAFWSGRFSLMDTGAKI
 HIPITIASVSHOPTWVSFFDFHLVCTFPAGLWFCIKINIDERVFLVALYISAVF
 AGVMVRLMLITPVVCMLSAIPSNVPEHVLGDMKRENPVEDSDSDKPNGLY
 DKAKVAKHTEQKREGLEGPNIKSIIVTMLMLLMFAVCTWTSNAYSPPSVL
 ASYHNDGTNLDDFRAIFWLNQNDDEHARVMSWDYGIQIAGMANRTLLVDNNWN
 NSHIALVGVKAVSSNETAAYKIMRSLDWDYVYLVIFGVGVIYSGDDINKFLWMVRIAGE

ORIGIN
 Query Match 83.8%; Score 2078.6; DB 3; Length 2709;
 Best Local Similarity 90.2%; Pred. No. 0;
 Matches 2238; Conservative 0; Mismatches 234; Indels 9; Gaps 1;
 polyA_signal /note="putative"
 polyA_site /note="putative"
 1 ATGCGGAGCCCTCGGCGCGGAGACAGACAGTGTGCTCCCTCAATCGTCCCGG 60
 76 ATGCGGAGCCCTCGGCGCGGAGACAGACAGTGTGCTCCCTCAATCGTCCCGG 135
 61 AGTGGCTCATGCGCCCTGGGAAACAGCCGCGACCAATGGCGCGCGGCGGCG 120
 136 AGCGGCTCATGCGCTCTGGGAAACAGCCGCGACCAATGGCGCGCGGCGGCG 195
 121 GCSCAAGGCGGCGGCGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGGCG 180
 196 CGTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 246
 181 GGGCTGTGCGAGCGCGCTGGGTGGCAGTGGCTTCTCTTCCATCCTCTCTG 240
 247 GGTGTGTGCGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 306
 241 TGGCTTGGCGGCTTACGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
 307 TGGCTGGCGGCTTACGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 366
 301 GAGTTCGACCGCGGTTAACTATAGATCAACACATCATCTTTCATGCGGTCT 360
 367 GAGTTCGACCGCGGTTAACTATAGATCAACACATCATCTTTCATGCGGTCT 426
 361 GAATTTTAAATTTGATGATGAAGAGCATGATATCCATAGGAGATAGTAGTGT 420
 427 GAGTTTCTAAATTTGATGATGAAGAGCATGATATCCATAGGAGATAGTAGT 486
 421 ACTGTTTACCCAGGTTGATGATAACCGGTGCGCTTATTCATTTGATTTAAAT 480
 487 ACCGTTTACCCAGGTTGATGATAACCGGTGCGCTTATTCATTTGATTTAAAT 546
 481 ACATTAACGTTTACATTAAGAGATGATGATGATGATGATGATGATGATGAT 540
 547 ACATTAACGTTTACATTAAGAGATGATGATGATGATGATGATGATGATGAT 606
 541 ACATTAACGTTTACATTAAGAGATGATGATGATGATGATGATGATGATGAT 600
 607 ACATTAACGTTTACATTAAGAGATGATGATGATGATGATGATGATGATGAT 666
 601 GCTGCTGTTTATTCGCTATGATGATGATGATGATGATGATGATGATGATGAT 660
 667 GCTGCTGTTTATTCGCTATGATGATGATGATGATGATGATGATGATGATGAT 726
 661 GATAATGAAGCGCATGCTATTTTTCGCTATGATGATGATGATGATGATGATGAT 720
 727 GATAATGAAGCGCATGCTATTTTTCGCTATGATGATGATGATGATGATGATGAT 786
 721 GTAAATACTGGGTCAGTTTTCGCTATGATGATGATGATGATGATGATGATGAT 780
 787 GTAAATACTGGGTCAGTTTTCGCTATGATGATGATGATGATGATGATGATGAT 846
 781 TCTGCTTGGGTCAGTTTTCGCTATGATGATGATGATGATGATGATGATGATGAT 840
 847 TCTGCTTGGGTCAGTTTTCGCTATGATGATGATGATGATGATGATGATGATGAT 906
 841 TTACTGATGAGAGATACAGCAAGAGCTACATAGCATATAGCACCTTCTACATT 900
 907 CTGCTGATGAGAGATACAGCAAGAGCTACATAGCATATAGCACCTTCTACATT 966

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 00:24:44 ; Search time 7543.8 seconds
(without alignments)
11984.275 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagccctcgccccc.....tatctagaagaacttttaa 2481

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*

2: gb_est2.*

3: gb_hcc.*

4: gb_est3.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_gss1.*

9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2078.6	83.8	2709	3 AK018758	Mus muscu
2	2055.6	82.9	2669	3 AK012153	Mus muscu
3	1640.6	66.1	4513	3 BC044321	Xenopus l
4	981.8	39.6	2334	3 AY310157	Rattus no
5	851	34.3	2212	3 CR610915	full-leng
6	809.8	32.6	1088	5 BX363211	EX363211
7	767.6	30.9	950	5 BU515954	AGENCOURT
8	731.2	29.5	852	7 CN458732	UI-M-HO0-
9	727	29.3	735	4 BG698273	602660188
10	698.4	28.1	879	5 EQ947308	AGENCOURT
11	695.8	28.0	823	7 CF736482	UI-M-HO0-
12	692.6	27.9	778	4 BG573592	602594608
13	686	27.6	867	4 B181585	603378796
14	685.2	27.6	861	7 CF995035	AGENCOURT
15	674.6	27.2	798	7 CN537592	UI-M-HO0-
16	667.4	26.9	3802	3 AK030363	Mus muscu
17	663.4	26.7	2656	3 BC028897	Mus muscu
18	658	26.5	2730	3 AK031547	Mus muscu
19	655.4	26.4	3643	3 AK077877	Mus muscu
20	649.6	26.2	754	7 CK636437	UI-M-HO0-
21	648.8	26.2	814	7 CN531243	UI-M-HO0-
22	640.2	25.8	766	7 CK633751	UI-M-HO0-
23	634	25.6	877	6 CB208902	AGENCOURT
24	633.4	25.5	757	4 B1869525	603392966

25	623.6	25.1	738	7	CN527774	UI-M-HO0-
26	621	25.0	759	4	B1652821	603286252
27	620.4	25.0	2103	9	AY418286	Mus muscu
28	618.6	24.9	891	5	EX669060	EX669060
29	618.6	24.9	2118	9	AY418284	Homo sapi
30	617.6	24.9	957	5	BU554374	AGENCOURT
31	616.4	24.8	776	4	BG620025	602618240
32	615	24.8	627	4	BG434322	602506307
33	614.8	24.8	825	6	CA318063	UI-M-FW0-
34	614	24.7	2727	3	AK087470	Mus muscu
35	613.4	24.7	709	7	CF780700	AGENCOURT
36	613	24.7	503	6	CD358477	AGENCOURT
37	611.8	24.7	779	5	BM963786	UI-M-EQ0-
38	611.8	24.7	779	6	CD351656	UI-M-G10-
39	611.4	24.6	775	7	CK032376	AGENCOURT
40	611	24.6	747	2	BE952197	UI-M-CD0-
41	611	24.6	2419	3	CNS0ADYZ	Arabidops
42	609.8	24.6	733	6	CA321357	UI-M-FW0-
43	609.6	24.6	689	2	BF970812	602273941
44	609.4	24.6	702	4	B1254237	602974710
45	609.4	24.6	724	7	CF773921	UI-M-HD0-

ALIGNMENTS

RESULT 1	AK018758	2709 bp	mRNA	linear	HTC 03-APR-2004
LOCUS	Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300006C19 product:hypothetical Oligosaccharyl transferase (OTase) STT3 subunit containing protein, full insert sequence.				
ACCESSION	AK018758				
VERSION	AK018758.2	GI:26384577			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1	Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Mus.			
AUTHORS	Garninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitesunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
MEDLINE	20530913				
PUBMED	11076861				
REFERENCE	4				
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.				
TITLE	Functional annotation of a full-length mouse cDNA collection				
JOURNAL	Nature 409, 685-690 (2001)				
REFERENCE	5				

1113	Db		GGATCCCTCTTATGCTAAGAAATAACAATCCCAATTATTCATCTGTTTCTGAGCACAGGC	1172
1224	Qy		TACGACTTGGGTGCTCTTCTCTTTGATCTACATATCTTGTATGTACCTTCCACAGG	1283
1173	Db		CACAACCTGGTCTCTCCCTACTATTTTGTATCTACAGCTCCTTGTCTTCATGTTCCAGTTGG	1232
1284	Qy		CCTTTGGTTCGTCATCAAAAATATCAAGTGAAGAGATTTCTGCTCTATATGCAT	1343
1233	Db		CCCTATTACTGCTTTAGCAACCTGCTGATGCTCGAATTTTATCATCATGATATGGTGT	1292
1344	Qy		CAGTCTGTCTACTTTGCTGGAGTGATGGTGGCACTGAATGTAGCTTTCAGTCCAGTCTGT	1403
1293	Db		GACCAGCATGTACTTTTTCAGCTGTAAATGTGTCGTCTAATGCTGGTATGCGACCTGTTAT	1352
1404	Qy		GTGTATCCTGCTGTGCAATTCCTTTTCAAATGCTTTTTCAGCACATTTTGGGGATGACAT	1463
1353	Db		GTGCATTCCTTTCTGGCAATGGTGTTCACGCT-----	1385
1464	Qy		GAAGAAGGAAATCCACCTGTGGAGCAGCAGTGATGAGATGACAAAAGAACCAAGG	1523
1386	Db		-----GCTGTCCACATATATCAAAA	1405
1524	Qy		AAATTTGTATGATAAGCAGGTAAAGTGAGGAAAATGCAACTCAACAGGAAAAAACTGA	1583
1406	Db		ATCTGGACATAAGTCGCCCAGACAAGAAGCAACAGGATTCCTACTTACCCCTA	1465
1584	Qy		AGAGGGATTAGGCCCTTAATATAAAGCATTTGTCCACCATGTTGATGCTGATGCTATGTAT	1643
1466	Db		TTAAGATGAGG-----TGCGAGTGGGATGATCTGCTCAGGCTTTTCTCTCAT	1517
1644	Qy		GATCTTTGCTGTCCACTGTACCTGGGTCAACAGCAANGCCTPACTCTAGTCCAAAGTGTAGT	1703
1518	Db		CACCTACACGCTTTCATTTCGACTTGGGTGACCAAGTCAAGCCTATTCTCTCCCTCCATGCT	1577
1704	Qy		CCTGGCCTACAAATCATGATGGCACCAAGGAATCTTATGATCATTTTAGAGAACTTA	1763
1578	Db		ACTGTCTGCTGTGGTGGGATGGCAGTAGGATCATTTTGTATGACTTCGAGAGCGTA	1637
1764	Qy		CTTTTGGCTAAGCGAAATACAGATGAACATGACGAGTAATGTCTTGGTGGGATTTAGG	1823
1638	Db		TTATTGGCTCCGTCACAATCTCCAGAGGATGCAAAAGTCATGTCTGTTGGGATTTAGG	1697
1824	Qy		CTATCAGATAGCTGGGAATGGCTAAATAGAACTACGTTGGTGATATAACACCTCGGAATAA	1883
1698	Db		CTACCAATTTACTTGCATGGCAATCGGACAATTTTAGTGACATTAACATGGAATAA	1757
1884	Qy		CAGCCACATAGCTGGTGGGAAAAGCTATGTCTTAATGAACACAGACCTTATAAAT	1943
1758	Db		TACCCATATTCTCGATAGGCGAGCAATGGCATCCACAGAGAAAAGCCCTATGAAT	1817
1944	Qy		CATGAGGACTTAGATGTAGATATGTTTGGTTATTTTGGAGGGGTATTGGCTATTTC	2003
1818	Db		CATGAGGAGCTTGATGCAGCTATGTGCTTCTCATTTTGGAGGCTTACTGGGTATTTC	1877
2004	Qy		TGGTGATGATATCAAGAAATTTCTCTGGATGGTAGGATAGCTCAAGG---AGAACATCC	2060
1878	Db		TTCCGATGATATCAACAGTTTCTTGGATGTCCGATTTGGAGAAAGCACAGACAGG	1937
2061	Qy		CAAGAATTCGGGAAAGTACTATTTTATCCCAAGGGGAATTCGCTGTACAAAAGC	2120
1938	Db		AAGACACATTAAGGAGAATGACTACTATCTCTACTGGGAAATCCGCTGTTGATCGTGA	1997
2121	Qy		AGGATCCCTACTTTGTGAATTCGCTTATGTATATAAATGTCTACTACAGATTTGGAGA	2180
1998	Db		GGGTCTCCGGTCTGCTCTCACTGCTCTTATGTACAAATGTGTACTACCGCTTTGGGCA	2057
2181	Qy		AATGCAGCTGGATTTTGTACACCCCGAGTTTTCAGCGAACAGTAAATGCTCAGATTGG	2240
2058	Db		GGTCTACACAGAACCCAGCGTCCACCGGCTTTTACCGTGTTCGAAATGCTCAGATTGG	2117
2241	Qy		AAATAAGGACATTAATAATTCAAACATTTGGAAGAGCCCTTTATCATCAGAACACTGGCTGT	2300


```
US-10-264-049-630
; Sequence 630, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAI33P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 630
; LENGTH: 2660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2634)..(2634)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2637)..(2637)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-630

Query Match 36.0%; Score 992; DB 16; Length 2660;
Best Local Similarity 99.8%; Pred. No. 3.5e-238;
Matches 892; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1598 GGATTAGGCCCTTAATATAAAGCATTTGCACCATGTTGATGCTGATGCTATTGATGAG 1647
DB 123 GGGTAGGCCCTTAATATAAAGCATTTGTCACCATGTTGATGCTGATGCTATTGATGAG 182
QY 1648 TTGCTGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1707
DB 183 TTGCTGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 242
QY 1708 GCCTCATACATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1767
DB 243 GCCTCATACATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
QY 1768 TGGCTAAGGCAAAATACAGATGAACGACGACGACGACGACGACGACGACGACGAC 1827
DB 303 TGGCTAAGGCAAAATACAGATGAACGACGACGACGACGACGACGACGACGACGAC 362
QY 1828 CAGATAGCTGGAATGGCTTAATAGAACTACGTTGGTGGATAATAACACCTGGAAACAC 1887
DB 363 CAGATAGCTGGAATGGCTTAATAGAACTACGTTGGTGGATAATAACACCTGGAAACAC 422
QY 1888 CACATAGCAGCTGGTGGGAAAGCTATGCTCTTAATGAACACGACGACCTATAAATCAT 1947
DB 423 CACATAGCAGCTGGTGGGAAAGCTATGCTCTTAATGAACACGACGACCTATAAATCAT 482
QY 1948 AGGACTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2007
DB 483 AGGACTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 542
QY 2008 GATGATATCAAAATTTCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 2067
DB 543 GATGATATCAAAATTTCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 602
QY 2068 ATTCGGGAAGTGACTATTTTACCCACAGGAGAAATTCGCTGTAGCAAAAGCAGGATCC 2127
DB 603 ATTCGGGAAGTGACTATTTTACCCACAGGAGAAATTCGCTGTAGCAAAAGCAGGATCC 662
QY 2128 CCTACTTTGTAATGCCCTATGCTATAAATGTGCTATGCTATGCTATGCTATGCTATG 2187
DB 663 CCTACTTTGTAATGCCCTATGCTATAAATGTGCTATGCTATGCTATGCTATGCTATG 722

US-10-296-115-629
; Sequence 629, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 629
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-629

Query Match 35.0%; Score 867.2; DB 16; Length 1114;
Best Local Similarity 98.8%; Pred. No. 1.7e-231;
Matches 895; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

QY 1 ATGCGGAGCCCTCGGCCCGCGGAGCAAGCAAGTGTGCTCAACTGCTCCCGTGG 60
DB 211 ATGCGGAGCCCTCGGCCCGCGGAGCAAGCAAGTGTGCTCAACTGCTCCCGTGG 270
QY 61 AGTGGCTCTATGCTGCTGGGACACGCGGACGCGGACGCGGCGCGGCGCGGCGCGGCG 120
DB 271 AGTGGCTCTATGCTGCTGGGACACGCGGACGCGGACGCGGCGCGGCGCGGCGCGGCG 330
QY 121 GCGCACAAAGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGG 179
DB 331 GCGCACAAAGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGG 390
QY 180 GGGGCTGCTGCGACGCGGCTGGTGGGAGTGGCTCTCTCTTCCATCCTCTCTCTCTCT 239
DB 391 GGGGCTGCTGCGACGCGGCTGGTGGGAGTGGCTCTCTCTTCCATCCTCTCTCTCTCT 450
QY 240 CTGGCTTGGCGGCTTCAGCTCGGCTCTTCGCGGCTCTTCGCGGCTCTTCGCGGCTCT 299
DB 451 CTGGCTTGGCGGCTTCAGCTCGGCTCTTCGCGGCTCTTCGCGGCTCTTCGCGGCTCT 510
QY 300 CGAGTTTCGACCGCTGGTAACTATAGATCAACATCATCTTGTGATCTCTGATCTCTG 359
DB 511 CGAGTTTCGACCGCTGGTAACTATAGATCAACATCATCTTGTGATCTCTGATCTCTG 570
QY 360 TGAATTTTAAATTTGTTTATGATGAAAGAGCTGATGATGATGATGATGATGATGATG 419
DB 571 TGAATTTTAAATTTGTTTATGATGAAAGAGCTGATGATGATGATGATGATGATGATG 630
```

QY 371 ATTGGTTTGAATGAAGAGCATGGTATCCACTAGAGAAATPAGTAGTGGTACTGTTTACC 430
 Db 282 ACTGGTTCGAGCGAGCGGATGATATCCGCTCGGAGGATGTGGCGGTACCGTCTATC 341
 QY 431 CAGGGTTGATGATACCGCTGGCTTATTCATTGGATTTTAAATACATTGACATACATG 490
 Db 342 CCGGCTGATGATGATACGTCGCGGGAATCCATTGGCTGTCGACAGTACTCAACATACCG 401
 QY 491 TTACATGAAGAGAGTATGTGTCTTCTGCACCAACTTTTAGCGCCTTACATCTATAT 550
 Db 402 TCCATATTTCGAGATCTGCGTGTCTGCGCGGATCTTCAGTGGCTGACCTCATCT 461
 QY 551 CTACTTTCCTGCTTACAGAGAACTTTGGAACCAAGAGCAGACATTTTAGCTGCTGTT 610
 Db 462 CCACCTACTCTGACCAAGAGAGTGTGGTCCGGGCGCGGCTTTCGCGCGCAGCT 521
 QY 611 TTATTGCTATTGTACCGGCTACATATCTCGGTGAGTGGATCTTTTGATTAATGAAG 670
 Db 522 TCATCGCCATCTGCTGGCTACATCAGTAGTGGTGGCTGGATCGTACGATAACGAGG 581
 QY 671 GCATTGCTATTTCGCACTTCAGTTCACATACATATTTATGGTAAATCTGTAAACATG 730
 Db 582 GCATTGCTATTTCGCACTTCAGTTCACATACATATTTATGGTAAATCTGTAAACATG 641
 QY 731 GGTCAAGTTTTTGGCAATGTGCTTATCTTATCTTATCTTATCTTATCTTATCTTAT 790
 Db 642 GATCGTGTCTGCTGGCGGAGCGCTTGTCTTACTTCTATCTATCTATCTATCTATCT 701
 QY 791 GTGGTTATGATTTATCATCAATCTTATTCACATGATGATTTGTTGTTACTGATGC 850
 Db 702 GTGGTTATGATTTATCATCAATCTTATTCACATGATGATTTGTTGTTACTGATGC 761
 QY 851 AGAGTACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATGTGGTATTAAT 910
 Db 762 GCAGTACTCTGCGCGTCTGCTGACAGTACAGCACTTCTACATCTGGGACTGCTGT 821
 QY 911 TATCAATCAGATACCTTTTGGGATTCAGCCCATCAGAACATGACATGACATGGAG 970
 Db 822 TCTCATGAGATCCCTTCTGGGATTCAGCCCATCAGAACATGACATGACATGGAG 881
 QY 971 CTGAGGTGTCTTTGCATGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGAGCCGAT 1030
 Db 882 CGCTGGAGTGTGTTGCTCTTATGCGCGTGGCACCTTTGCGCCATTTGCACTCGCTGC 941
 QY 1031 TAAACAAACAGATTCAGACCTTTCTTTTGGGTGATCAGTATCTGAGTGGCTG 1090
 Db 942 TGTCCGCAACAGATTCAGACCTTTCTTTTGGGTGATCAGTATCTGAGTGGCTG 1001
 QY 1091 TGTTCCTTAGTGTATCTATTGACTTTATACAGTATACATTTGACCATGAGTGGCAGGT 1150
 Db 1002 TCTTTGTCGCGTCTGCTGCTCACCATGCTGGCGTGTGGCGCGTGGAGTGGAGCT 1061
 QY 1151 TTTATTCATTGGGATCTGGGTATGCAAAATACATTTCCAAATTTGATGATGATG 1210
 Db 1062 TCTACTCGCTGGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1121
 QY 1211 CTGAGCATCAACCTACGATCTGGTGTCTTCTTCTTGTATCTACATATTTCTGTATGA 1270
 Db 1122 CGGAGCATCAGCCCACTTGGTGTCTGCTTCTTCTTGTATCTGACATCTGCTGCTG 1181
 QY 1271 CCTTCCAGAGCGCTTTGGTGTCTGATCAAAATATCAAGATGAAGATTTTGTG 1330
 Db 1182 CCTTCCAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1241
 QY 1331 CTCTATATGCAATCAGTGTCTCTTCTTCTTCTGAGTGTGCTGCTGCTGCTGCTGCT 1390
 Db 1242 TGTCTGATCGCATCAGTGTGCTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1301
 QY 1391 TGACCTCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1450
 Db 1302 TCACGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1361
 QY 1451 TGGGGGATGA-----CATGAAAGGGAATCCACCTGTGGAG-----GACAGCAGTG 1498

Db 1362 TGCAAGAGATTCGTCTAAGCGAATGGGCACAGCCATTAAGCGCACCCAGCGAGTGGATG 1421
 QY 1499 ATCAGGATGACAAAAGAAACCAAGGAATTTGTATGATAAGCGCAGGTAAAGTGAAGAAAC 1558
 Db 1422 AAGCTGAGATTCATTTGAGAGAGAGAGCGCTGTACGACAGGCTGGCAAGCTGAAGCATC 1481
 QY 1559 ATGCAACTGAAACAGGAAAACTGAAGAGAGGATTAAGCCCTTAATATAAAGACATTTGCA 1618
 Db 1482 GTACTAAGCATGATGCCAGCAGGATATCGGCGTCACTCCAACTGAAGAGTATTTGTTA 1541
 QY 1619 CCAATGTTGATGCTGATGCTATTGATGATGTTGCTGCTCCACTGCTGCTGCTGCTGCTG 1678
 Db 1542 TTTTGGCGCTTCTAATGCTGTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1601
 QY 1679 ATGCTACTCTAGTCCAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1735
 Db 1602 ATGCTACTCTAGTCCAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1661
 QY 1736 ATATCTTAGATGATTTTAGAGAGGCTTACTTTGGCTAAGGCAAAATACAGATGAACATG 1795
 Db 1662 ACATTTAGAGGATTTACAGAGGCTTACTACTGCTTTTCGAGACACACTCCCGATGATG 1721
 QY 1796 CAGAGTAAATGCTTGGTGGGATTTAGGCTATCAGATAGCTGGATGGCTTAATAGACTA 1855
 Db 1722 CTCGCGTATGCTTGGTGGGATTTAGGCTATCAGATAGCTGGATGGCTTAATAGACTA 1781
 QY 1856 CGTTGCTGATTAATAACCTCGAATAACAGCCACATAGCACTGCTGGTGGGAAAAAGCTATGT 1915
 Db 1782 CGCTAGTGGATTAATAACCTCGAATAACAGCCACATAGCACTGCTGGTGGGAAAAAGCTATGT 1841
 QY 1916 CTTCTTAATGAACAGCAGGCTTAATAATCATGAGAGCTCTAGATGTAGATTTATTTTGG 1975
 Db 1842 CTTCAACCGAGGAGAGTCTTACGAAATTTAGCATCTCTTGGAGTGGACTACGCTTTTGG 1901
 QY 1976 TTATTTTGGAGGGTATTGCTATTTCTGCTGATGATATCAACAAATTTCTCTGGATGG 2035
 Db 1902 TGATCTTTGGCGTGTATGCTGCTATTTCTGGGATGATATCAACAAATTTCTCTGGATGG 1961
 QY 2036 TTAGGATGCTGAAGAGAACTTCCAAAGACATTCGGGAAAGTGAATTTTACCCGAC 2095
 Db 1962 TCCGAAATTTCTGAGGAGAGTCTCCAAAGACATTTAAGGAAAGGATTTTACCCGAC 2021
 QY 2096 AGGAGAAATTTCTGCTGAGAAAGAGGATCCCTTCTTGTGATTTGCTTATGATATA 2155
 Db 2022 GCGTGAATTCAGGTTAGATGCGGAGTGTCTCCGCGCTGCTCACTGCTTATGTACA 2081
 QY 2156 AATGTCTATCTACAGATTTGGAGAAATGCACTGCTGATTTTCTGACCCCGAGTTTG 2215
 Db 2082 AATTAAGTACTACAGATTTGGGGAATTTGAAGTTTGGACTACAGAGTCCATCTGGATATG 2141
 QY 2216 ACCGACACGTTATGCTGAGATTTGGAATTAAGGACATTAATTTCAACATTTGGAAGAAG 2275
 Db 2142 ATCGACACGTTATGCTGAGATTTGGAATTAAGGACATTAATTTCAACATTTGGAAGAAG 2201
 QY 2276 CTTTACATCAGAACTGCTGCTTGTAGATTAATAAGTAAAGCAGCTGATTAACAGG 2335
 Db 2202 CCTACACGAGAACTGCTGCTTGTTCGCTATATAGGTTGAAGAGCGGATGAGTTCA 2261
 QY 2336 AGACATTAAGTACAAACCTCGAGTCCCAACATTTTCCCAAAACAGAAAGTATTGTTCAA 2395
 Db 2262 ATAGCCATCAGTGAAGCCAGGAGAGAACGATTTCTCCAGCAATTTCTATTTTCGAGAA 2321
 QY 2396 AGAAGACTACAAAGAGAGCGTGGCTCATTTAAATAAGCTGGTTTTTAAGAAAGCA 2455
 Db 2322 AGAATCTAAGCGTCCCAAG---GGCTACATACGAAACCGACCGGTTGTTGTTAAGGAA 2378
 QY 2456 AGAAATATCTAAGAGAC 2474
 Db 2379 AACGACCTTGAATAAAC 2397

; PRIOR APPLICATION NUMBER: PCT/US98/23435
 ; PRIOR FILING DATE: 1998-11-04
 ; PRIOR APPLICATION NUMBER: US 60/064,911
 ; PRIOR FILING DATE: 1997-11-07
 ; PRIOR APPLICATION NUMBER: US 60/064,912
 ; PRIOR FILING DATE: 1997-11-07
 ; PRIOR APPLICATION NUMBER: US 60/064,983
 ; PRIOR FILING DATE: 1997-11-07
 ; PRIOR APPLICATION NUMBER: US 60/064,900
 ; PRIOR FILING DATE: 1997-11-07
 ; PRIOR APPLICATION NUMBER: US 60/064,988
 ; PRIOR FILING DATE: 1997-11-07
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 611
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 133
 ; LENGTH: 1543
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1055)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1143)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; US-10-621-401-133

 Query Match 42.7%; Score 1060.4; DB 16; Length 1543;
 Best Local Similarity 99.8%; Pred. No. 1.6e-285;
 Matches 1061; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 267 CTTGCGCTGATCGCTTCGAAACATCCAGAGTTCGACCCGCTGGTTAACTATAG 326
 Db 1 CTTGCGCTGATCGCTTCGAAACATCCAGAGTTCGACCCGCTGGTTAACTATAG 60

 QY 327 ATCAACACATCATCTTGATCTCATCTGAGTTCTATGAATTTTAAATTTGTTGATGAAG 386
 Db 61 ATCAACACATCATCTTGATCTCATCTGAGTTCTATGAATTTTAAATTTGTTGATGAAG 120

 QY 387 AGCATGGTATCCATGAGAGAAATAGTAGTGGTACTGTTTACCCAGGTTGATGATAC 446
 Db 121 AGCATGGTATCCATGAGAGAAATAGTAGTGGTACTGTTTACCCAGGTTGATGATAC 180

 QY 447 CGCTGGCCTTATTCATTGGATTTTAAATAATTGAACATAACTGTTTACATAAGAGACT 506
 Db 181 CGCTGGCCTTATTCATTGGATTTTAAATAATTGAACATAACTGTTTACATAAGAGACT 240

 QY 507 ATGTGTGTTCTTCGACCACTTTTAGCGCCTTACATCTATCTACTTCTCTGCTTAC 566
 Db 241 ATGTGTGTTCTTCGACCACTTTTAGCGCCTTACATCTATCTACTTCTCTGCTTAC 300

 QY 567 AAGAGAACTTTGGAACCAAGAGAGAGACTTTTAGCTGCTGTTTATTGCTATTGTACC 626
 Db 301 AAGAGAACTTTGGAACCAAGAGAGAGACTTTTAGCTGCTGTTTATTGCTATTGTACC 360

 QY 627 AGGCTACATATCTCGTTCAGTGGTCTGATCTCTTTGATTAATGAAGGATTCATTTTGC 686
 Db 361 AGGCTACATATCTCGTTCAGTGGTCTGATCTCTTTGATTAATGAAGGATTCATTTTGC 420

 QY 687 ACTTCAGTTCACATCTATTTATGGGTAAATCTGTAAAACTCGGTTCAGTTTGGAC 746
 Db 421 ACTTCAGTTCACATCTATTTATGGGTAAATCTGTAAAACTCGGTTCAGTTTGGAC 480

 QY 747 AATGTGCTGCTTATCTCTATTTCTATATGGTCTCTGCTGGGTGGTTATGATTTAT 806
 Db 481 AATGTGCTGCTTATCTCTATTTCTATATGGTCTCTGCTGGGTGGTTATGATTTAT 540

 QY 807 CATCAATCTTATTCATGCAATGATTTTGTGTTGTTTACTGATCCAGAGATACAGCAAAAG 866
 Db 541 CATCAATCTTATTCATGCAATGATTTTGTGTTGTTTACTGATCCAGAGATACAGCAAAAG 600

QY 867 AGCTACATAGCATATAGCACTTTCTACATTTCTGGTTTATATATATATCAATCGCATACC 926
 Db 601 AGCTACATAGCATATAGCACTTTCTACATTTCTGGTTTATATATATATCAATCGCATACC 660

 QY 927 TTTTGTGGGATTCAGCCCAATCAGAACAAAGTGAACACATGCGAGCTGCGAGTGTCTTTGC 986
 Db 661 TTTTGTGGGATTCAGCCCAATCAGAACAAAGTGAACACATGCGAGCTGCGAGTGTCTTTGC 720

 QY 987 ATTGCTGCAAGCTTATGCTTTCTGAGTATCTGAGAGACGATTAACAAAACAGAGCTT 1046
 Db 721 ATTGCTGCAAGCTTATGCTTTCTGAGTATCTGAGAGACGATTAACAAAACAGAGCTT 780

 QY 1047 CCAGACCTTTCTTTTGGGTATCTACTAGCTGAGTGTCTGTTTCTTAGTGTAT 1106
 Db 781 CCAGACCTTTCTTTTGGGTATCTACTAGCTGAGTGTCTGTTTCTTAGTGTAT 840

 QY 1107 CTATTTGACTTATACAGGTTTACATTCACCAGTGGAGTGGAGTTTATTCATTGTGGGA 1166
 Db 841 CTATTTGACTTATACAGGTTTACATTCACCAGTGGAGTGGAGTTTATTCATTGTGGGA 900

 QY 1167 TACTGGGTATGCAAAAATACATTCCAATTTATGATCATGAGTGTGAGCATCAACCTAC 1226
 Db 901 TACTGGGTATGCAAAAATACATTCCAATTTATGATCATGAGTGTGAGCATCAACCTAC 960

 QY 1227 GACTTGGGTGCTTTCTTTGATCTACATATTTCTTGTATGATCTTCCAGAGGCT 1286
 Db 961 GACTTGGGTGCTTTCTTTGATCTACATATTTCTTGTATGATCTTCCAGAGGCT 1020

 QY 1287 TTGGTTCTGCATCAAAAATATCAAGATGAAGAGATTTGTT 1329
 Db 1021 TTGGTTCTGCATCAAAAATATCAAGATGAAGAGATTTGTT 1063

RESULT 11
 US-10-028-384-7
 ; Sequence 7, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREAU, Claude
 ; APPLICANT: MCBRIDE, Kevin
 ; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
 ; FILE REFERENCE: 5600-74
 ; CURRENT APPLICATION NUMBER: US/10/028,384
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 7
 ; LENGTH: 2417
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: AF132552
 ; DATABASE ENTRY DATE: 1999-04-27
 ; RELEVANT RESIDUES: (1) .. (2417)
 ; US-10-028-384-7

Query Match 38.5%; Score 955.8; DB 15; Length 2417;
 Best Local Similarity 64.8%; Pred. No. 4.4e-256;
 Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;

QY 191 AGCGGCTGGGTGGAGTCTCTCTCTTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 250
 Db 102 AGGTGGCTGGTACAGCAGCCTAATCACCTCGCCATCTCTCTCTCTCTCTCTCTCTCTCT 161

 QY 251 GCTTCAGCTCGCGCTCTTCGCGGTCATCCGCTTCGAAAGCATCATCCAGAGTTCCAGC 310
 Db 162 GATTTCT 221

 QY 311 CGTGGTTAACTATAGATCAACACATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 370
 Db 222 CGTGGTTAACTATAGATCAACACATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 281

387 AGCATGGTATCCACTAGGAAGATAGTAGGGTGGTACTGTTTACCCAGGGTTGATGATAC 446
Db AGCATGGTATCCACTAGGAAGATAGTAGGGTGGTACTGTTTACCCAGGGTTGATGATAC 180
QY 447 CGCTGGCTTATTCATTTGGATTTTAAATACATTTGAACATAAATGTTTACATTAAGAGAGCT 506
Db CGCTGGCTTATTCATTTGGATTTTAAATACATTTGAACATAAATGTTTACATTAAGAGAGCT 240
QY 507 ATGTGTGTTCTTGGACCAACTTTAGGGCTTACATCTATATCTACTTCTCTGCTTAC 566
Db ATGTGTGTTCTTGGACCAACTTTAGGGCTTACATCTATATCTACTTCTCTGCTTAC 300
QY 567 AAGAGAACTTTGGAACCAAGAGCAGGACTTTAGCTGCTGTTTATTTATGCTATTGCTAC 626
Db AAGAGAACTTTGGAACCAAGAGCAGGACTTTAGCTGCTGTTTATTTATGCTATTGCTAC 360
QY 627 AGGCTACATATCTGGTCAGTACGATCTTTGATATTAAGAGGATGCTATTGCTTAC 686
Db AGGCTACATATCTGGTCAGTACGATCTTTGATATTAAGAGGATGCTATTGCTTAC 420
QY 687 ACTTCAGTTCACATATCTATTTATGGTAAATCTGTAATAAATCTGTAATAAATCTGTAATAA 746
Db ACTTCAGTTCACATATCTATTTATGGTAAATCTGTAATAAATCTGTAATAAATCTGTAATAA 480
QY 747 AATGTGCTGCTTATCTATTTCTATATGCTCTGCTTGGGGTGGTATGATTTAT 806
Db AATGTGCTGCTTATCTATTTCTATATGCTCTGCTTGGGGTGGTATGATTTAT 540
QY 807 CATCAATCTTATTCAGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 866
Db CATCAATCTTATTCAGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 867 AGTCTACATATCTAGGCTTCTACATGCTGCTTAAATATCTATGATGATGATGATGATGAT 926
Db AGTCTACATATCTAGGCTTCTACATGCTGCTTAAATATCTATGATGATGATGATGATGAT 660
QY 927 TTTTGGGATTCAGCAATCAGAACAGTGAACATGCGAGCTGCGAGGCTGCTTGGC 986
Db TTTTGGGATTCAGCAATCAGAACAGTGAACATGCGAGCTGCGAGGCTGCTTGGC 720
QY 987 ATTCTGCAAGCTTATGCTTCTGCTAGTATCTGAGACCGATTAACAAACAGAGTT 1046
Db ATTCTGCAAGCTTATGCTTCTGCTAGTATCTGAGACCGATTAACAAACAGAGTT 780
QY 1047 CCAGACCTTTTCTTTTGGGTGATCACTAGCTGCGAGGCTGCTTCTTATGATGAT 1106
Db CCAGACCTTTTCTTTTGGGTGATCACTAGCTGCGAGGCTGCTTCTTATGATGAT 840
QY 1107 CTATTTGATATACAGGTTACATTCAGCATGAGTGGCAGGTTTATTCATTTGGGA 1166
Db CTATTTGATATACAGGTTACATTCAGCATGAGTGGCAGGTTTATTCATTTGGGA 900
QY 1167 TACTGGGTATGCAAAATATACATTCATTTGATGCTGAGCATCAACCTAC 1226
Db TACTGGGTATGCAAAATATACATTCATTTGATGCTGAGCATCAACCTAC 960
QY 1227 GACTTGGGTGCTTCTTCTGATCTGATATCTGATGATGATGATGATGATGATGATGATGAT 1286
Db GACTTGGGTGCTTCTTCTGATCTGATATCTGATGATGATGATGATGATGATGATGATGAT 960
QY 1287 TTGGTCTGCTCAAAATATCAACGATGAAGAGATTTTGGT 1329
Db TTGGTCTGCTCAAAATATCAACGATGAAGAGATTTTGGT 1063

RESULT 8

US-09-818-683-133

; Sequence 133, Application US/09818683

; Publication No. US20030211472A1

; GENERAL INFORMATION:

; APPLICANT: Peng et al.

; TITLE OF INVENTION: 125 Human Secreted Proteins

FILE REFERENCE: P2020P1
CURRENT APPLICATION NUMBER: US/09/818,683
CURRENT FILING DATE: 2001-03-28
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 612
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 133
LENGTH: 1543

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1055)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1143)
OTHER INFORMATION: n equals a,t,g, or c
US-09-818-683-133

Query Match 42.7%; Score 1060.4; DB 10; Length 1543;
Best Local Similarity 99.8%; Pred. No. 1.6e-285;
Matches 1061; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 267 CTTCCGGCTCATCCGCTTCGAAAGCATCATCCAGAGTTCGACCGGTGTTTAACTATAG 326
Db 1 CTTCCGGCTCATCCGCTTCGAAAGCATCATCCAGAGTTCGACCGGTGTTTAACTATAG 60
QY 327 ATCAACACATCATCTTGCATCTCATGGGTTCATGAATTTTAAATTCGTTTGAAGAAG 386
Db 61 ATCAACACATCATCTTGCATCTCATGGGTTCATGAATTTTAAATTCGTTTGAAGAAG 120
QY 387 AGCATGGTATCCACTAGGAAGATAGTAGGGTGGTACTGTTTACCCAGGGTTGATGATAC 446
Db 121 AGCATGGTATCCACTAGGAAGATAGTAGGGTGGTACTGTTTACCCAGGGTTGATGATAC 180
QY 447 CGCTGGCTTATTCATTTGGATTTTAAATACATTAACATTAACATGTTTCAATAAGAGAGCT 506
Db 181 CGCTGGCTTATTCATTTGGATTTTAAATACATTAACATGTTTCAATAAGAGAGCT 240
QY 507 ATGTGTGTTCTTGGACCAACTTTTAGCGGCTTACATCTATATCTACTTTCCTGCTTAC 566
Db 241 ATGTGTGTTCTTGGACCAACTTTTAGCGGCTTACATCTATATCTACTTTCCTGCTTAC 300
QY 567 AAGAGAACTTTGGAACCAAGAGCAGGACTTTAGCTGCTGTTTATTTATGCTATTGCTAC 626
Db 301 AAGAGAACTTTGGAACCAAGAGCAGGACTTTAGCTGCTGTTTATTTATGCTATTGCTAC 360
QY 627 AGCTACATATCTCGGTGAGTAGCTGGATCCCTTTGATTAATGAAGCATGCTATTGCTG 686
Db 361 AGCTACATATCTCGGTGAGTAGCTGGATCCCTTTGATTAATGAAGCATGCTATTGCTG 420
QY 687 ACTTCAGTTCACATATCTATTTATGGTAAATCTGTAATAAATCTGTAATAAATCTGTAATAA 746
Db 421 ACTTCAGTTCACATATCTATTTATGGTAAATCTGTAATAAATCTGTAATAAATCTGTAATAA 480
QY 747 AATGTGCTGCTGCTTATCTATTTCTATATGCTCTGCTTGGGGTGGTATGATTTAT 806
Db 481 AATGTGCTGCTGCTTATCTATTTCTATATGCTCTGCTTGGGGTGGTATGATTTAT 540
QY 807 CATCAATCTTATTCAGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 866
Db 541 CATCAATCTTATTCAGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 867 AGTCTACATATAGGCTTCTACATGCTGCTTAAATATCTATGATGATGATGATGATGAT 926
Db 601 AGTCTACATATAGGCTTCTACATGCTGCTTAAATATCTATGATGATGATGATGATGAT 660
QY 927 TTTTGGGATTCAGCAATCAGAACAGTGAACATGCGAGCTGCGAGGCTGCTTGGC 986
Db 661 TTTTGGGATTCAGCAATCAGAACAGTGAACATGCGAGCTGCGAGGCTGCTTGGC 720
QY 987 ATTCTGCAAGCTTATGCTTCTGCTAGTATCTGAGACCGATTAACAAACAGAGTT 1046
Db 721 ATTCTGCAAGCTTATGCTTCTGCTAGTATCTGAGACCGATTAACAAACAGAGTT 1046

US-09-974-879-133

Query Match 42.7%; Score 1060.4; DB 10; Length 1543;
Best Local Similarity 99.8%; Pred. No. 1.6e-285;
Matches 1061; Conservative 0; Mismatches 2; Indels 0;

Qy	267	CTTCCGCGTCATCCGCTTCGAAAGCATCATCCACGAGTTTCGACCCGTCGGTTTAACTATAG	326
Db	1	CTTCCGCGTCATCCGCTTCGAAAGCATCATCCACGAGTTTCGACCCGTCGGTTTAACTATAG	60
Qy	327	ATCAACACATCATCTTGCACTCATGCGTTCATGAATTTTTAAATTCGTTTGCATGAAG	386
Db	61	ATCAACACATCATCTTGCACTCATGCGTTCATGAATTTTTAAATTCGTTTGCATGAAG	120
Qy	387	AGCATGGTATCCACTAGGAAGAATPAGTAGTGTACTGTTTACCAGGGTTGATGATAAC	446
Db	121	AGCATGGTATCCACTAGGAAGAATPAGTAGTGTACTGTTTACCAGGGTTGATGATAAC	180
Qy	447	CSCTCGCCCTTATTCATTCGGATTTAAATACATTTGAACATAACTGTTTCACATAAGACAGCT	506
Db	181	CSCTCGCCCTTATTCATTCGGATTTAAATACATTTGAACATAACTGTTTCACATAAGACAGCT	240
Qy	507	ATGTGTGTTCCCTTGCAACCACTTTAGCGCCCTTACATATATCTACTTTCCCTGCTTAC	566
Db	241	ATGTGTGTTCCCTTGCAACCACTTTAGCGCCCTTACATATATCTACTTTCCCTGCTTAC	300
Qy	567	AGAAGAATTGGACACAGGACGAGCACTTTTAGCTGCTGTTTATTGCTATTGTACC	626
Db	301	AGAAGAATTGGACACAGGACGAGCACTTTTAGCTGCTGTTTATTGCTATTGTACC	360
Qy	627	AGGCTACATATCTCGGTCAGTAGCTGGATCCTTTGTAATGAAGCAATGCTATTTTGC	686
Db	361	AGGCTACATATCTCGGTCAGTAGCTGGATCCTTTGTAATGAAGCAATGCTATTTTGC	420
Qy	687	ACTTCAGTTCACATACATTTATCGGTAAATCTGTAAATACTGGGTGAGTTTTTTCGAC	746
Db	421	ACTTCAGTTCACATACATTTATCGGTAAATCTGTAAATACTGGGTGAGTTTTTTCGAC	480
Qy	747	AAATGCTGCTGCTTATCCTATTTCTATATGTCCTCTGCTGGGTGGTATGATATTAT	806
Db	481	AAATGCTGCTGCTTATCCTATTTCTATATGTCCTCTGCTGGGTGGTATGATATTAT	540
Qy	807	CATCAATCTTATCCACTGCATGATTTGTTGTTTACTGATGCAGAGATACAGCAAG	866
Db	541	CATCAATCTTATCCACTGCATGATTTGTTGTTTACTGATGCAGAGATACAGCAAG	600
Qy	867	AGTCTACATAGCATATAGCACTTTCTACATTTGGGTTTTAAATATTATCAATGCAGATACC	926
Db	601	AGTCTACATAGCATATAGCACTTTCTACATTTGGGTTTTAAATATTATCAATGCAGATACC	660
Qy	927	TTTTGTGGGATCCAGCCAAATCAGAAACAGTGAACATGGCAGCTGCAGGTGCTTTGC	986
Db	661	TTTTGTGGGATCCAGCCAAATCAGAAACAGTGAACATGGCAGCTGCAGGTGCTTTGC	720
Qy	987	ATTGCTGAAGCTTATGCTTTCTTTGCAGTATCTGCAGAGCCGATTAACAAAACAGAGTT	1046
Db	721	ATTGCTGAAGCTTATGCTTTCTTTGCAGTATCTGCAGAGCCGATTAACAAAACAGAGTT	780
Qy	1047	CCAGACCCCTTTCTTTTGGGTGATACCTAGCTGCAGGTGCTGTGTTCTTAGTGTCAT	1106
Db	781	CCAGACCCCTTTCTTTTGGGTGATACCTAGCTGCAGGTGCTGTGTTCTTAGTGTCAT	840
Qy	1107	CTATTTGACTTATACAGGTTACATATGCACCATGGATGGCAGGTTTTATTCAATTTGGGA	1166
Db	841	CTATTTGACTTATACAGGTTACATATGCACCATGGATGGCAGGTTTTATTCAATTTGGGA	900
Qy	1167	TACTGGGATGCAAAAATACACATTCCAATATTGTCATCAGTGTCTGACCATCAACCTAC	1226
Db	901	TACTGGGATGCAAAAATACACATTCCAATATTGTCATCAGTGTCTGACCATCAACCTAC	960
Qy	1227	GACTTGGGTGTCCTTTCTTTTGATCTACATATCTTTGATGTACATCTTCCAGCAGGCGCT	1286
Db	961	GACTTGGGTGTCCTTTCTTTTGATCTACATCTTTGATGTACATCTTCCAGCAGGCGCT	1020

QY	1287	TTGGTTC	TGCA	TCAAAAA	TATCA	ACGATGA	AAAGCTATTT	TGTT	1329
DB	1021	TTGGTTC	TGCA	TCAAAAA	TATCA	ACGATGA	AAAGANTATTT	TGTT	1063

RESULT 7

```

US-09-305-736-133
; Sequence 133, Application US/09305736
; Publication No. US20030089078A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/305,736
; CURRENT FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: PCT/US98/23435
; EARLIER FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/064,911
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,912
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,983
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,900
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,988
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,987
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,908
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,984
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,985
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/066,094
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,100
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,089
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,095
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,090
; EARLIER FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-305-736-133

```

Query Match	42.7%;	Score 1060.4;	DB 10;	Length 1543;
Best Local Similarity	99.8%;	Pred. No. 1.6e-285;		
Matches 1061;	Conservative	0;	Mismatches 2;	Indels 0; Gaps 0;

267	QY	CTTGGCGGTCATCGCGCTTGGAAAGCATCATCGACGAGTTCGACCCCGTGGTTAACTCTATAG	326
1	Db	CTTGGCGGTCATCGCGCTTGGAAAGCATCATCGACGAGTTCGACCCCGTGGTTAACTCTATAG	60
327	QY	ATCAACACATCATCTTCGATCTCATGGGTTCTATGAATTTTAAATTTGGTTTGATGCAAG	386
61	Db	ATCAACACATCATCTTCGATCTCATGGGTTCTATGAATTTTAAATTTGGTTTGATGCAAG	120

```

; SEQ ID NO 412
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1203)..(1203)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-237-412

Query Match      42.8%; Score 1061; DB 16; Length 1209;
Best Local Similarity 99.8%; Pred. No. 9.5e-286;
Matches 1061; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 267 CTTGCCGCTCATCGCTTCGAAAGCATCATCCACGAGTTCGACCGTGTGTTTAACTATAG 326
DB 1 CTTGCCGCTCATCGCTTCGAAAGCATCATCCACGAGTTCGACCGTGTGTTTAACTATAG 60
QY 327 ATCAACACATCATCTTGCATCTCATGGTCTCTATGAATTTTAAATTTGGTTTGAAG 386
DB 61 ATCAACACATCATCTTGCATCTCATGGTCTCTATGAATTTTAAATTTGGTTTGAAG 120
QY 387 AGCATGTATCCACTAGGAAGATAGTAGGTGTACTGTTTACCAGGTTGATGATAAC 446
DB 121 AGCATGTATCCACTAGGAAGATAGTAGGTGTACTGTTTACCAGGTTGATGATAAC 180
QY 447 CGTGGCCCTTATTCATGATGATCTCATGGTCTCTATGAATTTTAAATTTGGTTTGAAG 506
DB 181 CGTGGCCCTTATTCATGATGATCTCATGGTCTCTATGAATTTTAAATTTGGTTTGAAG 240
QY 507 ATGTGTCTTCTTGCACCACTTTTAGCGCCCTTACATCTATCTACTTCTCTCTTAC 566
DB 241 ATGTGTCTTCTTGCACCACTTTTAGCGCCCTTACATCTATCTACTTCTCTCTTAC 300
QY 567 AAGAGAACTTTGGAACCAAGGACGAGCTTTTAGCTGCTGTTTATGCTATTTGAC 626
DB 301 AAGAGAACTTTGGAACCAAGGACGAGCTTTTAGCTGCTGTTTATGCTATTTGAC 360
QY 627 AGGTACATATCTCGGTCAGTCTGATCTGATGATGATGATGATGATGATGATGAT 686
DB 361 AGGTACATATCTCGGTCAGTCTGATCTGATGATGATGATGATGATGATGATGAT 420
QY 687 ACTTCAGTTCACATATCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 746
DB 421 ACTTCAGTTCACATATCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 747 AATGTGCTGCTGCTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 806
DB 481 AATGTGCTGCTGCTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 807 CATCAATCTTATTCCTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 866
DB 541 CATCAATCTTATTCCTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 867 AGTCTACATAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 926
DB 501 AGTCTACATAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 927 TTTTGTGGGATTTCCAGCCATCAGAACAGTGAACATGCGAGTGCAGGTGCTTTTGC 986
DB 661 TTTTGTGGGATTTCCAGCCATCAGAACAGTGAACATGCGAGTGCAGGTGCTTTTGC 720
QY 987 ATTGCTGCAAGCTTATGCTTCTTCTGAGTATCTGAGAGACCGATTAACAAACAGAGTT 1046
DB 721 ATTGCTGCAAGCTTATGCTTCTTCTGAGTATCTGAGAGACCGATTAACAAACAGAGTT 780
QY 1047 CCAGACCCCTTTCTTTTGGGTGTATCATAGCTGAGGTGCTGCTTCTTCTTCTTCTTCT 1106
DB 781 CCAGACCCCTTTCTTTTGGGTGTATCATAGCTGAGGTGCTGCTTCTTCTTCTTCTTCT 840
QY 1107 CTATTTGACATTATACAGGTTTACATGTCAGGAGTGCAGGTTTATTCATCTGCGGA 1166
DB 841 CTATTTGACATTATACAGGTTTACATGTCAGGAGTGCAGGTTTATTCATCTGCGGA 900

```

```

QY 1167 TACTGGGTATGCAAAATAACACATTCACATTCATTCATTCATTCATTCATTCATTCAT 1226
DB 901 TACTGGGTATGCAAAATAACACATTCACATTCATTCATTCATTCATTCATTCATTCAT 960
QY 1227 GACTTGGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1286
DB 961 GACTTGGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1020
QY 1287 TTGGTTCTGCATCAAAATAATCAACGATGAAGAGTATTTGTT 1329
DB 1021 TTGGTTCTGCATCAAAATAATCAACGATGAAGAGTATTTGTT 1063

RESULT 6
US-09-974-879-133
; Sequence 133, Application US/09974879
; Publication No. US20030028003A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P2
; CURRENT APPLICATION NUMBER: US/09/974,879
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/239,893
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/918,683
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 09/305,736
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: PCT/US98/23435
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/064,911
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,912
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,983
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,900
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,988
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,987
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,908
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,984
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,985
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/066,094
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,100
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,089
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,095
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,090
; PRIOR FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c

```


QY 2126 CCTACTTTGTTGAATTCCTTATGTATATAAATGTCTATCATCTACAGATTGGAGAAATGCG 2187
 Db 784 CCTACTTTGTTGAATTCCTTATGTATATAAATGTCTATCATCTACAGATTGGAGAAATGCG 843
 QY 2188 CTGGATTTTCGTACACCCCCAGGTTTGGACCGAACAACGTAATGCTGAGATTGGAAATAAG 2247
 Db 844 CTGGATTTTCGTACACCCCCAGGTTTGGACCGAACAACGTAATGCTGAGATTGGAAATAAG 903
 QY 2248 GACATTAAATTCACAAATTTGGAAGAGCCTTTACATCAGACACCTGCTGTTAGGATA 2307
 Db 904 GACATTAAATTCACAAATTTGGAAGAGCCTTTACATCAGACACCTGCTGTTAGGATA 963
 QY 2308 TATAAAGTAAAGCACTGATACAGGAGACATTAGATCAAAACCTCGAGTCACCAAC 2367
 Db 964 TATAAAGTAAAGCACTGATACAGGAGACATTAGATCAAAACCTCGAGTCACCAAC 1023
 QY 2368 ATTTCCTCCAAACAGAGTATTTGTCAGAGAGACTTACCAAGAGCGTGGCTACATT 2427
 Db 1024 ATTTCCTCCAAACAGAGTATTTGTCAGAGAGACTTACCAAGAGCGTGGCTACATT 1083
 QY 2428 AAAAATAAGCTGGTTTTTAAAGAAAGCAAGAAATATCTAAGAGACTGTTTTAA 2481
 Db 1084 AAAAATAAGCTGGTTTTTAAAGAAAGCAAGAAATATCTAAGAGACTGTTTTAA 1137

RESULT 4
 US-10-106-698-330
 ; Sequence 330, Application US/10106698
 ; Publication No. US20030109690A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
 ; FILE REFERENCE: PA005P1
 ; CURRENT APPLICATION NUMBER: US/10/106,698
 ; CURRENT FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524
 ; PRIOR FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: US 60/157,137
 ; PRIOR FILING DATE: 1999-09-29
 ; PRIOR APPLICATION NUMBER: US 60/163,280
 ; PRIOR FILING DATE: 1999-11-03
 ; NUMBER OF SEQ ID NOS: 8564
 ; SOFTWARE: Patent In Ver. 3.0
 ; SEQ ID NO 330
 ; LENGTH: 1209
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1203)..(1203)
 ; OTHER INFORMATION: n equals a,t,g, or c

Query Match 42.8%; Score 1061; DB 15; Length 1209;
 Best Local Similarity 99.8%; Pred. No. 9.5e-286;
 Matches 1061; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 267 CTTCCGCGTCATCCGCTTCGAAAGCATATCCACGAGTTCCGCGTGGTTAACTATAG 326
 Db 1 CTTCCGCGTCATCCGCTTCGAAAGCATATCCACGAGTTCCGCGTGGTTAACTATAG 60
 QY 327 ATCAACACATCATCTTGCATCTCATGGTTCTATGAATTTTAAATTTGGTTGATGAAG 386
 Db 61 ATCAACACATCATCTTGCATCTCATGGTTCTATGAATTTTAAATTTGGTTGATGAAG 120
 QY 387 AGCATGTATCCACTAGGAAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATAAC 446
 Db 121 AGCATGTATCCACTAGGAAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATAAC 180
 QY 447 CGCTGGCTTATTCATGGATTTTAAATACATTGAACATACTGTTACATAAGAGCGT 506
 Db 181 CGCTGGCTTATTCATGGATTTTAAATACATTGAACATACTGTTACATAAGAGCGT 240

QY 507 ATGTGTGTTCTCTTGACCAACCTTTTAGCGGCTTACATCTATATCTACTTTCTCTGCTTAC 566
 Db 241 ATGTGTGTTCTCTTGACCAACCTTTTAGCGGCTTACATCTATATCTACTTTCTCTGCTTAC 300
 QY 567 AAGAGAACTTTGGAAACCAAGAGAGAGACTTTTACGTGCTTGTGTTTATGCTATGTGACC 626
 Db 301 AAGAGAACTTTGGAAACCAAGAGAGAGACTTTTACGTGCTTGTGTTTATGCTATGTGACC 360
 QY 627 AGGCTACATATCTCCGTGAGTAGCTGATCCTTTGATAAAGAGAGCAATTCGATTTTTCG 686
 Db 361 AGGCTACATATCTCCGTGAGTAGCTGATCCTTTGATAAAGAGAGCAATTCGATTTTTCG 420
 QY 687 ACTTCAGTTTCACATCTACTTTTATGGGTAAATCTGTAAAAAATCTGGGTGCTGTTTTGGAC 746
 Db 421 ACTTCAGTTTCACATCTACTTTTATGGGTAAATCTGTAAAAAATCTGGGTGCTGTTTTGGAC 480
 QY 747 AATGTGCTGCTGCTTATCCTATTTCTATATGCTCTGCTTGGGTGCTGTTATGTTAT 806
 Db 481 AATGTGCTGCTGCTTATCCTATTTCTATATGCTCTGCTTGGGTGCTGTTATGTTAT 540
 QY 807 CATCAATCTTATTCACCTGCTGATTTTGTGTTTCTGTTACTGATCAGAGATACAGCAAAAG 866
 Db 541 CATCAATCTTATTCACCTGCTGATTTTGTGTTTCTGTTACTGATCAGAGATACAGCAAAAG 600
 QY 867 AGCTACATAGCATATAGCACTTTCTACATTTGCTGGTTTAAATATTAATCAATCGAGATACC 926
 Db 601 AGCTACATAGCATATAGCACTTTCTACATTTGCTGGTTTAAATATTAATCAATCGAGATACC 660
 QY 927 TTTTGTGGGATTCAGCCCAATCAGAAAGAGTGAACACATGCGAGCTCGAGGTGCTTTGTC 986
 Db 661 TTTTGTGGGATTCAGCCCAATCAGAAAGAGTGAACACATGCGAGCTCGAGGTGCTTTGTC 720
 QY 987 AATGTGCAAGCTTATGCTTTCTGAGTATCTGAGAGCCGATTACAAAACAGAGTT 1046
 Db 721 AATGTGCAAGCTTATGCTTTCTGAGTATCTGAGAGCCGATTACAAAACAGAGTT 780
 QY 1047 CCAGACCCCTTTCTTTTGGGTGATCAGTAGCTGCGAGGTGCTGTTGCTTGTAGTGTGAT 1106
 Db 781 CCAGACCCCTTTCTTTTGGGTGATCAGTAGCTGCGAGGTGCTGTTGCTTGTAGTGTGAT 840
 QY 1107 CTATTGTACTTATACAGTTTACATTGACAGTGGAGTGGAGGTTTATTCATTGTTGGGA 1166
 Db 841 CTATTGTACTTATACAGTTTACATTGACAGTGGAGTGGAGGTTTATTCATTGTTGGGA 900
 QY 1167 TACTGGGTATGCAAAAATACACATTTCCAAATTTATTCATGCTGCTGAGCATCAACCTAC 1226
 Db 901 TACTGGGTATGCAAAAATACACATTTCCAAATTTATTCATGCTGCTGAGCATCAACCTAC 960
 QY 1227 GACTTGGGTGCTTTCTTTTGTGATCTACATATTTCTTGTATGTAACCTTCCAGCAGGCT 1286
 Db 961 GACTTGGGTGCTTTCTTTTGTGATCTACATATTTCTTGTATGTAACCTTCCAGCAGGCT 1020
 QY 1287 TTGGTTCTGCTCAAAAATATCAACGATGAAAGAGTATTTGTT 1329
 Db 1021 TTGGTTCTGCTCAAAAATATCAACGATGAAAGAGTATTTGTT 1063

RESULT 5
 US-10-264-237-412
 ; Sequence 412, Application US/10264237
 ; Publication No. US20040009491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birse et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PA131P1
 ; CURRENT APPLICATION NUMBER: US/10/264,237
 ; CURRENT FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450
 ; PRIOR FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: US 60/205,515
 ; PRIOR FILING DATE: 2000-05-19
 ; NUMBER OF SEQ ID NOS: 2876
 ; SOFTWARE: Patent In Ver. 3.1

Db	2281	ACATCAGAACACTGGCTTGTGTAGGATATATPAAAGTAAAGCACCTGTATACAGGGAGACA	2340
QY	2341	TTAGATCACAAACCTCGAGTCCACCACTTTCCAAAAAGAAGTATTTGTCAAAGAG	2400
Db	2341	TTAGATCACAAACCTCGAGTCCACCACTTTCCAAAAAGAAGTATTTGTCAAAGAG	2400
QY	2401	ACTCCCAAAGGAGCGTGCTACATTAATTAATTAAGCTGGTTTTTAAGAAAGGCAAGAAA	2460
Db	2401	ACTCCCAAAGGAGCGTGCTACATTAATTAATTAAGCTGGTTTTTAAGAAAGGCAAGAAA	2460
QY	2461	ATATCTAAGAGACTGTTTAA	2481
Db	2461	ATATCTAAGAGACTGTTTAA	2481

RESULT. 2

```

US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication NO. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK019758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)..(2469)
US-10-028-384-3

```

Query	Match	83.8%;	Score	2078.6;	DB	15;	Length	2710;	
Best	Local Similarity	90.2%;	Pred.	No. 0;					
Matches	2238;	Conservative	0;	Mismatches	234;	Indels	9;	Gaps	1;
QY	1	ATGCGGAGCCCTCGGCCCGGAGACAAAGCAAGTCTCCCTCAACTGCTCCCGGTGG	60						
Db	77	ATGCGGAGCCCTCGGCCCGGAGACAAAGCAAGTCTCCCTCAACTGCTCCCGGTGG	136						
QY	61	ATGTGCCCTCATGGCCCTGGGAAACAGCCGGCAGGCAACCAAGGSCCGCGGGCCCACTGC	120						
Db	137	AGCGCCTCATGGCTCTGGGACAGCGCCACGGCCACCATGGCCCGGACCCAGAGC	196						
QY	121	CGGCACAAGCGCGCGCGCGCGCGCGCGCGAGCGCGGCCCGCGGGCGTGTCCGGG	180						
Db	197	GGCTCCAGGGG-----GGGCGCGGAAGCGGGGCCCCCGCGGGCGTGTCCGGG	247						
QY	181	GGCGTGTGCGAGCGCGTGGTGGCAGTCCGCTTCTCTCTTCAACATCTCTTCTCGGC	240						
Db	248	GGCTTGTGCGAGCGCGCGGGTGGCAGTGGTGTCTCTTCAACATCTCTTCTCGGC	307						
QY	241	TGGCTTGGCGGCTTCAGCTCGCGCTCTTTCGGCGTCAATCCGTTGCGAAGCATCATCCAC	300						
Db	308	TGGTGGCGGCTTCAGCTCGCGCTCTTTCGGCGTCAATCCGTTGAGAGCATCATCCAC	367						
QY	301	GAGTTCGACCGTGGTTTACTATATATCAACACATCATCTTGATCTCATGGTCTAT	360						
Db	358	GAGTTCGACCGTGGTTTAACTATATATCAACACATCATCTTGATCTCATGGTCTAT	427						
QY	361	GAATTTTAAATTCGTTTGATGAAGAGCATGGTATCCACTAGGAAGATAGTAGTGGT	420						
Db	428	GAGTTCTCAAAATTCGTTTGATGAAGAGCATGGTATCCCACTGGGAAGAAATAGTGGTGGC	487						
QY	421	ACTGTTTACCAGGGTTGATGATAACCGGTGCCCTTATTTCATTGGATTTTAAATACATTC	480						

	Db	489	ACCGTTTACCAGGGTTGATGAACAAGCTGGCCTATTTCATGTGAATTTAAATACATTG	547
	Qy	481	AACATAACTGTTACATAAAGAGACGTATGTGTGTCTTGTCACCAACTTTTTAGCGGCCTT	540
	Db	548	AACATAACAGTTACATAAAGAGATGTGTGTATCTCTTGCCACCAACTTTTAGCGGCCTT	607
	Qy	541	ACAICTATATCTACTTTCCTGCTTAACAAGAGAACITTTGGAACCAAGSAGAGACTTTTA	600
	Db	608	ACATCCATATCTAGCTTCCCTGCTAACTTAGAGAACTGTGGACCAAGSAGAGACTTCTA	667
	Qy	601	GCTGCTGTGTTTTATTGCTATTATACAGGCTACATATCTCGGTCAGTAGCTGATCCCTT	660
	Db	668	GCTGCTGCTTCATTGCTATCGTTACCAGGGTACATATCTCGGTCAGTGCGGATCCTTT	727
	Qy	661	GATATGAGGCCATTGCTATTTTGTGCACITTCAGTTACATATCATATTTATGGTAAAATCT	720
	Db	728	GATATGAGGCATTGCCATTTTGGCTTCAGTTCACTTACTCTTAAGGTAAAGTCT	787
	Qy	721	GTAAAACTGGGTCAGTTTTTGGCAAAATGTCTGCTGCTATCTCTATTTCATATGCTC	780
	Db	788	GTGAAGACCGGCTCTGTGTCTTGCAAAATGTCTGCTGCTGTCACTATTTCTACATGGTC	847
	Qy	781	TCTGCTTGGGGTGTATGTATTTATCATCAATCTATTCCACTGSCATGATTTGTGCTG	840
	Db	848	TCTGCTGGGAGGTTATGTGTTTCAATCAACCTCACTCCCTCCCATGTGTTGTGTTG	907
	Qy	841	TTACTGATGCAGATACACAAAAGAGTCTACATAGCATATAGCACHTTCTCAATTTG	900
	Db	908	CTGTGTATGCAGAGGTACAGCAAGAGAGTCTACATAGCATATAGCACHTTGTACATTTG	967
	Qy	901	GSTTTAATATTATCAATGCAGATACCTTTTGTGGGATTCAGGCCAATCAGAAACAAGTAA	960
	Db	968	GSTTTAATATTATCCAATGCAGATACCTTTTGTGGGATTCAGGCCAATCAGAAACAAGCAG	1027
	Qy	961	CACATGGCAGCTGCAGGTGCTTTTGCAATGTCTGCAAGCTTATGCTTTCTCGCATCTG	1020
	Db	1028	CACATGGCAGCTGCAGGTGCTTTTGCGCTGCTGCAAGCTTACGCTTTTTTGCAGTATCTG	1087
	Qy	1021	AGAGACCGATTAAACAAAACAAGAGTTCCAGAGCCCTTTCTTTTGGGTGATCACTAGCT	1080
	Db	1088	AGAGACCGTTTGACAAAACAAGAGTTCCAGAGCCCTTTCTTTTGGGTGATCACTAGCT	1147
	Qy	1081	GCAGTGCTGTGTTTCTTATGTGTCATCTATTTTGAAGCTATACAGGTTACANTGCACATGG	1140
	Db	1148	GCAGCGCTGTGTTTCTTATGTGTCATCTATCTGACATACAGAGTTATATTGCACCATGG	1207
	Qy	1141	AGTGGCAGTTTTATTCAATTGGGATACGGGTATGCAAAATACACATTCCAATTAAT	1200
	Db	1208	AGTGGCAGTTTTATTCAATTGGGATACGGGTATGCAAAATACACATTCCAATTAAT	1267
	Qy	1201	GCATCAGTGCTGAGCATCAACCTAGCATCTGGGTGCTTTCTCTTTTGATCTCATATT	1260
	Db	1268	GCATCAGTGCTGAGCATCAGCCTAGCATCGGGTGCTTTCTCTTTTGATCTCATATT	1327
	Qy	1261	CTTGATGTACTCTCCAGCAGGCCCTTGTTCTGTCATCAAATAATCAACGATGAAGA	1320
	Db	1328	CTTGATGTACTCTCCAGCAGGCCCTATGTTCTGTCATCAAATAATCAACGATGAAGA	1387
	Qy	1321	GTATTTGTGCTCTATATGCATCAGTGCTGTCTTCTCTTTTGATCTCATATT	1380
	Db	1388	GTATTTGTGCTCTGTATGCGATCAGTGCTGTGTACTTTGCGAGGTGANGGTGCGGCTG	1447
	Qy	1381	ATGTTGACTTTGATCTCCAGTCGTGTGTATGCTGTCTGCAATTGCTTTTCAAATGTTTTT	1440
	Db	1448	ATGCTGACTCTGACCCCGGTCTCTGTCATGCTGTGCGCCATCGCTCTTCCAAATGTTTT	1507
	Qy	1441	GAGCACTATTTGGGGNATGACATGAAGAGGAAAATCCACCTGTGGAGGACAGCAGTGAT	1500
	Db	1508	GAGCACTATTTGGGGATGACATGAAGAGGAAAATCCCACTGTGTGGAGGACAGCAGTGAT	1567
	Qy	1501	GAGGATGACAAAAGAAAACCAAGGAAATTTGTATGATGAAGGACGGTAAAGTGAGGAAACAT	1560

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 07:01:53 ; Search time 1239.99 Seconds
(without alignments)
11031.546 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagacccctggcccc.....tatctaagaagactgtttaa 2481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4085731 seqs, 2755760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2481	100.0	2481	US-10-028-384-1	Sequence 1, Appli
2	2078.6	83.8	2710	US-10-028-384-3	Sequence 3, Appli
3	1114.8	44.9	1828	US-09-945-527-62	Sequence 62, Appl
4	1061	42.8	1209	US-10-106-698-330	Sequence 330, App
5	1061	42.8	1209	US-10-264-237-412	Sequence 412, App
6	1060.4	42.7	1543	US-09-974-879-133	Sequence 133, App
7	1060.4	42.7	1543	US-09-305-736-133	Sequence 133, App
8	1060.4	42.7	1543	US-09-818-683-133	Sequence 133, App
9	1060.4	42.7	1543	US-09-818-683-133	Sequence 133, App
10	955.8	38.5	2417	US-10-621-401-133	Sequence 133, App
11	955.8	38.5	2417	US-10-028-384-7	Sequence 7, Appli
12	892	36.0	2660	US-10-264-049-630	Sequence 630, App

13	867.2	35.0	1114	16	US-10-236-115-629	Sequence 629, App
14	689.2	27.8	2256	15	US-10-032-585-6323	Sequence 6323, Ap
15	670.6	27.0	3093	18	US-10-417-375-92	Sequence 92, Appl
16	670.6	27.0	3094	15	US-10-028-384-9	Sequence 9, Appli
17	646.6	26.1	2472	15	US-10-171-581-112	Sequence 112, App
18	646.6	26.1	2472	15	US-10-028-384-11	Sequence 11, Appl
19	646.6	26.1	2472	15	US-10-172-118-742	Sequence 742, App
20	646.6	26.1	2472	16	US-10-342-887-742	Sequence 742, App
21	646.6	26.1	3046	18	US-10-417-375-95	Sequence 95, Appl
22	638	25.7	5404	18	US-10-417-375-99	Sequence 99, Appl
23	638	25.7	5404	18	US-10-417-375-97	Sequence 97, Appl
24	616.2	24.8	2839	18	US-10-425-115-150745	Sequence 150745
25	610.2	24.6	2881	18	US-10-425-114-14408	Sequence 14408, A
26	609.4	24.6	2866	16	US-10-320-797-2305	Sequence 2305, Ap
27	606.2	24.4	2232	15	US-10-128-714-7139	Sequence 7139, Ap
28	585.2	23.6	2779	17	US-10-437-963-99904	Sequence 99904, A
29	539.4	21.7	2157	9	US-09-801-368-387	Sequence 387, App
30	539.4	21.7	2157	18	US-10-793-639-318	Sequence 318, App
31	537.8	21.7	2733	15	US-10-028-384-5	Sequence 5, Appli
32	515	20.8	1848	15	US-10-128-714-2139	Sequence 2139, Ap
33	500	20.2	500	9	US-09-988-588-1643	Sequence 1643, Ap
34	458.2	18.5	558	14	US-10-052-283-433	Sequence 433, App
35	456.4	18.4	1969	15	US-10-128-714-1139	Sequence 1139, Ap
36	456.4	18.4	2603	15	US-10-128-714-6139	Sequence 6139, Ap
37	456.4	18.4	3969	15	US-10-128-714-139	Sequence 139, App
38	456.4	18.4	4603	15	US-10-128-714-5139	Sequence 5139, Ap
39	450.4	18.2	3141	18	US-10-425-115-130787	Sequence 130787
40	443.4	17.9	485	10	US-09-918-995-11283	Sequence 11283, A
41	430	17.3	430	9	US-09-878-178-717	Sequence 717, App
42	430	17.3	430	13	US-10-046-935-717	Sequence 717, App
43	430	17.3	430	14	US-10-146-502-717	Sequence 717, App
44	429	17.3	3197	16	US-10-424-599-111541	Sequence 111541
45	415.4	16.7	616	9	US-09-879-536-332	Sequence 332, App

ALIGNMENTS

RESULT 1
US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREULT, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)
; OTHER INFORMATION:
US-10-028-384-1

Query Match	100.0%	Score	2481	DB	15	Length	2481
Best Local Similarity	100.0%	Pred. No.	0				
Matches	2481	Conservative	0	Mismatches	0	Indels	0
						Gaps	0
QY	1	ATGGCGGAGCCCTGGCGCCCGGAGAGCAAGCAAGTCCTCACTCGTCCCGTGG	60				
Db	1	ATGGCGGAGCCCTGGCGCCCGGAGAGCAAGTCCTCACTCGTCCCGTGG	60				
QY	61	AGTGGCCCTCATGGCCCTGGGAAAACAGCCGGCAGCCGACCGGGGCCCCAGTGC	120				
Db	61	AGTGGCCCTCATGGCCCTGGGAAAACAGCCGGCAGCCGACCGGGGCCCCAGTGC	120				

; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 21090
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-21090

Query Match 9.8%; Score 242.4; DB 4; Length 245;
Best Local Similarity 99.6%; Pred. No. 1.6e-55;
Matches 243; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	551	CTACTTTCCTCTTACAGAACTTTGGAAACCAAGGAGCAGACCTTTAGCTGCTGTT	610
DB	1	CTACTTTCCTCTTACAGAACTTTGGAAACCAAGGAGCAGACCTTTAGCTGCTGTT	60
QY	611	TTATGCTATTGTACCAAGGCTACATATCTCGGTCAGTAGCTGATCTTTGATAATGAAG	670
DB	61	TTATGCTATTGTACCAAGGCTACATATCTCGGTCAGTAGCTGATCTTTGATAATGAAG	120
QY	671	GCATTGCTATTGTCAGCTTCAGTTCACATCTATTATGGGTAAAAATCTGTAAAAAATG	730
DB	121	GCATTGCTATTGTCAGCTTCAGTTCACATCTATTATGGGTAAAAATCTGTAAAAAATG	180
QY	731	GGTCAAGTTTTTGGACAAATGCTGCTGCTTATACCTATTCTATATGGTCTCTGCTGG	790
DB	181	GGTCAAGTTTTTGGACAAATGCTGCTGCTTATACCTATTCTATATGGTCTCTGCTGG	240
QY	791	GTGG 794	
DB	241	GTGG 244	

RESULT 13

US-09-270-767-12856
; Sequence 12856, Application US/09270767
; Patent No. 6703491

GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12856

LENGTH: 914

TYPE: DNA

ORGANISM: *Drosophila melanogaster*

US-09-270-767-12856

Query Match 9.6%; Score 237.4; DB 4; Length 914;
Best Local Similarity 64.2%; Pred. No. 7.9e-54;
Matches 374; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY	1741	TTAGTATGATTTTACAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACCA	1800
DB	2	TTAGTATGATTTTACAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACCA	61
QY	1801	GTAATGCTCTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG	1860
DB	62	ATAATGCTCTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG	121
QY	1861	GTGGATATAACACCTGGATTAACAGGACATAGCACTGTGGGAAAGCTATGCTTCTT	1920
DB	122	GTGGATATAACACCTGGATTAACAGGACATAGCACTGTGGGAAAGCTATGCTTCTT	181
QY	1921	AATGAAACAGCAGGCTTATAAATCATGAGGACTCTAGATGATGATTTTGGTTAT	1980
DB	182	TCGGAGGAGAGAGCTACGAGATTAATGAGGAACTGATGCTGACTACGTTCTCGTAT	241
QY	1981	TTTGGAGGGTATTTGGCTATCTCGTGATGATTAACAAATTTCTCTGGATGGTTAG	2040

DB	242	TTCCGAGGGCTCACTGGCTACTCATCGGACGATATCAACAGTTCTCTGTGATGGTGGC	301
QY	2041	ATAGCTGAAGGAGAAATCCCAAG---ACATTGGGAAAGTGAATTTTACCACCAG	2097
DB	302	ATTGGCGGACGACGGATCGTGGTGGCACATCCGGAAGGACTACTATCGCGCCAC	361
QY	2098	GGAGAAATCCGTGTAGAACAAAGCAGATCCCTACTCTTTGTAANTGCTTATGATATA	2157
DB	362	GGAGAGTTCCGAGTGGACAAAGAGGGCTCACCCACATGCTCAATTTGTTGATACAG	421
QY	2158	ATGTCACTACTACAGATTTGGAGAAATGCGAGCTGGATTTTCGTACACCCCGAGTTTTC	2217
DB	422	ATGTGCTACTATCGCTTTGGGCAATGTACACGAGGTGGCAAGCCAGGCTACCAT	481
QY	2218	CGAACAGCTAATCTGAGATTGGAATTAAGGACATTAATTCACACATTTGGAAGAGCC	2277
DB	482	CGAGTTCTGTCGCCCGAGATCGGCAACAAAGGACTTTGAACTGGATGCTCTCGGAGGG	541
QY	2278	TTTACATCAGAACACTGGCTTGTTAGGATATATAAGTAAAG 2320	
DB	542	TACACCAAGGAGCACTGGCTGGTGGCACTACAGGTTAAG 584	

RESULT 14

US-09-270-767-1318/c
; Sequence 1318, Application US/09270767
; Patent No. 6703491

GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1318

LENGTH: 560

TYPE: DNA

ORGANISM: *Drosophila melanogaster*

US-09-270-767-1318

Query Match 8.1%; Score 201.4; DB 4; Length 560;
Best Local Similarity 63.9%; Pred. No. 2.9e-44;
Matches 342; Conservative 0; Mismatches 181; Indels 12; Gaps 2;

QY	1196	TTATTGCATCAGTGTCTGAGCATCACTACAGCTGGGTGCTTCTTCTTTGATCTAC	1255
DB	560	TCATTGCATCCGTGCGGAGCATCCACCACACTTGGTTCTGTTCTTTGATCTAC	501
QY	1256	ATATTCTTGTATGATCTTCCCAGCAGGCTTTGGTTCTGCAATCAAAATATCAACGATG	1315
DB	500	ACATCATGTTGTCGCTTCCCAGTGGAGTGTGGTACTGTCATCAAGCAGATCAACGACG	441
QY	1316	AAAGAGTATTTGTTGCTCTATATGCAATCAGTGTCTCTACTTCTGAGGATGGTGC	1375
DB	440	AGCGGTTTCTGTTGGTGTGTACGCCATCAGTGGGTTTACTTCTGTTGTTGTTGCT	381
QY	1376	GACTGATGTTGACTTTTCACTCCAGTCTGTGTATGCTGTCTGCAATTCCTTTTCAAA	1435
DB	380	GTTTGTATTTGACCTTCACTCCAGTGGTGTGTGATGCTGCGCGAGTGGGCTTTTCGG	321
QY	1436	TTTTTGAGCACTATTTGGGGGATGA-----CATGAAAGGGAATCCCACTGTGGAG	1488
DB	320	TGTTGATGTTCTTCCTGCAAGGATTCGTTAAGCAATGGGACAGCATTAACGCGAG	261
QY	1489	-----GACAGCAGTGTAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGAG	1543
DB	260	CCACCGAAGTGGATGAAGCTGAGGATTCATTTGAGAAGAGAGCGCTGTACGCAAGGCTG	201
QY	1544	GTAAGTGAGAAACATGCACTGACAGGAAAAAACTGAAGAGGATTAGGCCCTATA	1603
DB	200	GCAAGCTGAAGCATCTGTTACTTAAGCATGATGCCCGACAGGATACTGGCTGCTCAACC	141

Query Match	11.0%; Score 272.2; DB 4; Length 867;
Best Local Similarity	63.5%; Pred. No. 3.2e-63;
Matches	434; Conservative 0; Mismatches 243; Indels 6; Gaps 1;
1648	TTTGCTGTCCACTGTACCTGGGTGTCACCAAGCAATGCCCTACTCTAGTCCAAAGTGTAGTCG 1707
85	TTTGTTTTACATGTACTTTGGGTAAACATCGAATGCTTATTCATCACCACTCAGTGTGTTTA 144
1708	GCTCATACAATCATGATGGCACCAGGAATATCTTAGATGATTTTAGAGAGAGCTACTTT 1767
145	GCATCCAGAAACCCAGATGGCTCACAACTATCATTTGATGATATAGAGAGCCATTATAC 204
1768	TGGCTAAGCCAAATACAGATGAACATGCACGAGTAATGTCTTTGGTGGGATTATGGCTAT 1827
205	TGGTTAAGAATGAATACACCAAGAATGCCAAAGTTATGGCCCTGGTGGGATTTATGTTAT 264
1828	CAGATAGCTGGAAATGCTTAATAGAATACGTGGTGGATAATAACACCTTGGAAATACACG 1887
265	CAAAATCGGGGTATGCTGNTAGAACCAACTTTGTGTAAACATACATGATGATATACACA 324
1888	CACATAGCACTGGTGGGAAAAGCTATGTCTTCTAATGAACACAGCAGCCTATAAAATCATG 1947
325	CATATTCGCCACTGTTCGTGAAGCAATGTCTTCCCTGGAAGATGTGTGCTATGAAATTTG 384
1948	AGGACTCTAGATGATAGTATGTTTTGGTTATTTTTGGAGGGGTTATTCGGCTATTCGTG 2007
385	AGACACACGATGTTGATATGTGTAGTTATATTTGGAGGGTTATTTGGGTTATTCGTG 444
2008	GATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAGAGAAACATCCCAAGAC 2067
445	GATGATATTAAACAAATTCCTTAGGATGGAAGAANTGCTGAAGGTATCTTGGCCTGATGAA 504
2068	ATTCCGGAAGTGACTATTTTACCCACACAGGGAGAAATCCGTGTAGACAAAGCAGGATCC 2127
505	ATCAAAAGAAAGACGACTACTTTACTGACCAGGAGAAATATAAGTGGATAAAGATGCATCA 564
2128	CCCTACTTGTGTGAATTCGCTATGTATATAAATGCTCATCTACAGATTTGGAGAAATGCAG 2187
565	CTGGCAATGAAGAAATCTTTGATGTATAGTTATCGTNTCATAGATTCAGTNA----- 618
2188	CTGGATTTTCGTACACCCCAAGCTTTTGAACCAACAGTAATGCTGAGATGTGAATAGAG 2247
619	TTGTTTGGAGGTAGAGATGGTGTGTAGTAGAGTTAGAAACCAACAAATCCACAGCAATGAA 678
2248	GACATTAATAATTCAAACATTTGGAAGAAGCCTTTACATCAGAACACTGGCTTGTAGGATA 2307

RESULT 12
US-09-513-999C-21090
; Sequence 21090, Application US/09513999C
; Patent No. 5783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 5783961
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26

QY 1089 TGTGTTCTTGTAGTGTCTATTTTGGACTTATACAGG 1124
DB 1346 GGTGTTCTGTTGGTACCTCTGCTCAGGCTTACCGG 1381

RESULT 8

US-09-270-767-158/c
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-158

Query Match 11.0%; Score 273.8; DB 4; Length 1660;
Best Local Similarity 63.8%; Pred. No. 1.8e-63;
Matches 432; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
QY 1798 CGAGTAATGCTTGGTGGGATTTGGCTATCAGATAGCTGGAAATGGCTTAATAGAACTACG 1857
DB 829 CCAGTAGTAAGCCTCTCTGAAATCGTCTAAATGTTGCGGAAATGCCAAACAGACGACG 770
QY 1858 TTGTTGGATATACACCTGGAAATACAGCCACATAGCACTGTGTGGAAAGCTATGCT 1917
DB 769 CTAGTGGATAATAATAGCTGAACAATAGTCAATAGCGTGTGTGGCAGGCAATGCT 710
QY 1918 TCTAATGAACAGCAGCCTCTAAATCATGAGGACTCTAGATAGATTATGTTTGGTT 1977
DB 709 TCAACCGAGGAGAGTCTCTAGAAATATGACATCTCTTCACTGGACTACGTTTTGGTG 650
QY 1978 ATTTTGGAGGGTTATGGCTATCTGGTATGATATCAACAATTTCTGGATGTT 2037
DB 649 ATCTTTGGCGGTGATCGGCTATTCGGCGATGATCAACAAGTTCTCTGGATGGTC 590
QY 2038 AGGATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAG 2097
DB 589 CGAATTCGGAGGAGAGCATCCCAAGACATTAAGGAAAGCATTTACTTTACCGACCG 530
QY 2098 GGAGAAATTCGCTAGACAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGTATAA 2157
DB 529 GGTGAATTCAGGTAGATGCCAAGGTGCTCCGGCCCTGCTCACTGCTTATGTACAA 470
QY 2158 ATGTCATACTACGATTGGAGAAATGCAGCTGGAATTTGCTACACCCCGAGTTTGAC 2217
DB 469 TTAAGCTACTACGATTCGGGAAATGAAGTTGAGCTACAGAGGCCATCTGGATATGAT 410
QY 2218 CGACAGTATGCTGAGATTGGAATAGGACATTAATTCAAATTTGGAGAGCC 2277
DB 409 CGCACAGTAAACCGGCTCAITGGGAATAGGACTTTGATCTGACCTACCTGGAGAGGCC 350
QY 2278 TTTACATCAGAACACTGGCTTTGTTAGGATATATAAAGTAAAGCACTGATACAGGGAG 2337
DB 349 TACACCACAGAACACTGGCTTTGTCATCTATAGTGTGAAGAGCGGATGTTCAAT 290
QY 2338 ACATTAGATCACAACTCGAGTCACCAATTTCCCAAAACAGAGTATTTGTCAGAG 2397
DB 289 AGACATCAGTGAAGACAAGAGAGAGAACGATTCCTCCAGCA ---AATTCATCTCGAGA 233
QY 2398 AAGACTACCAAAAGAGCGGTGGCTACATTTAAAAATAAGCTGGTTTTTAAGAAAGCGAAG 2457
DB 232 AAGAACTCGAAGCGTCGCAAGGGCTACATACGAAACCGACCGGTTGTTGTTAAGGAAAA 173
QY 2458 AAAATATCTAGAGAGAC 2474

DB 172 CGAACCTTGAATAAAC 156

RESULT 9

US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-15440

Query Match 11.0%; Score 273.8; DB 4; Length 1660;
Best Local Similarity 63.8%; Pred. No. 1.8e-63;
Matches 432; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
QY 1798 CGAGTAATGCTTGGTGGGATTTGGCTATCAGATAGCTGGAAATGGCTTAATAGAACTACG 1857
DB 829 CCAGTAGTAAGCCTCTCTGAAATCGTCTAAATGTTGCGGAAATGCCAAACAGACGACG 770
QY 1858 TTGTTGGATATTAACACCTGGAAATACAGCCACATAGCACTGTGTGGAAAGCTATGCT 1917
DB 769 CTAGTGGATTAATAATAGCTGGAACAATAGTCAATAGCGTGTGTGGCAGGCAATGCT 710
QY 1918 TCTAATGAACAGCAGCCTCTAAATCATGAGGACTCTAGATAGATTATGTTTGGTT 1977
DB 709 TCAACCGAGGAGAGTCTCTAGAAATATGACATCTCTTCACTGGACTACGTTTTGGTG 650
QY 1978 ATTTTGGAGGGTTATGGCTATCTGGTATGATATCAACAATTTCTCTGGATGTT 2037
DB 649 ATCTTTGGCGGTGATCGGCTATTCGGCGATGATCAACAAGTTCTCTGGATGGTC 590
QY 2038 AGGATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAG 2097
DB 589 CGAATTCGGAGGAGAGCATCCCAAGGACATTAAGGAAAGCATTTACTTTACCGACCG 530
QY 2098 GGAGAAATTCGCTAGACAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGTATAA 2157
DB 529 GGTGAATTCAGGTAGATGCCAAGGTGCTCCGGCCCTGCTCACTGCTTATGTACAA 470
QY 2158 ATGTCATACTACGATTGGAGAAATGCAGCTGGAATTTGCTACACCCCGAGTTTGAC 2217
DB 469 TTAAGCTACTACGATTCGGGAAATGAAGTTGAGCTACAGAGGCCATCTGGATATGAT 410
QY 2218 CGAACAGTAACTCTGAGATTGGAATAAGGACATTAATTCAAACATTTGGAGAGGCC 2277
DB 409 CGCACAGTAAACCGGCTCAITGGGAATAGGACTTTGATCTGACCTACCTGGAGAGGCC 350
QY 2278 TTTACATCAGAACACTGGCTTTGTTAGGATATATAAAGTAAAGCACTGATACAGGGAG 2337
DB 349 TACACCACAGAACACTGGCTTTGTCATCTATAGTGTGAAGAGCGGATGTTCAAT 290
QY 2338 ACATTAGATCACAACTCGAGTCACCAATTTCCCAAAACAGAGTATTTGTCAGAG 2397
DB 289 AGACATCAGTGAAGACAAGAGAGAGAACGATTCCTCCAGCA ---AATTCATCTCGAGA 233
QY 2398 AAGACTACCAAAAGAGCGGTGGCTACATTTAAAAATAAGCTGGTTTTTAAGAAAGCGAAG 2457
DB 232 AAGAACTCGAAGCGTCGCAAGGGCTACATACGAAACCGACCGGTTGTTGTTAAGGAAAA 173
QY 2458 AAAATATCTAGAGAGAC 2474
DB 172 CGAACCTTGAATAAAC 156

```

; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Query Match      13.7%; Score 340.2; DB 4; Length 1660;
Best Local Similarity 67.0%; Pred. No. 2.2e-81;
Matches 499; Conservative 0; Mismatches 243; Indels 3; Gaps 1;

QY 1730 CCAGGATATCTTAGATGATTTAGAGAGCTTACCTTTGGCTTAAGGCAAAATACAGATG 1789
DB 789 CCCGCAACATTTAGACGATTTAGAGAGGCTTACTACTGGCTTTGCGAGACACTGCGG 848
QY 1790 AACATGACAGATGATGCTTGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATA 1849
DB 849 ATGATGCTCGGCTTATGCTTGTGGGATTTACGGATACAGATAGCGGGAATGGCAACA 908
QY 1850 GAACCTACGCTTGGTGGGATTAACACCTGGGAATTAACAGCCACATAGCACTGGTGGGAAAG 1909
DB 909 GAACGACGCTAGTGGATTAATAATACGTGGACATAGTACATAGGCTGGTTGGCAGG 968
QY 1910 CTATGCTCTTAATGAACAGAGGCTTATAAATCAATAGAGGACTTAGATGTAGATTAATG 1969
DB 969 CAATGCTCTTCAACCGGAGAGAGCTTACGAAATTTATGACATCTCTTGACGTGGACTAG 1028
QY 1970 TTTTGGTATTTTGGAGGGCTTATTCGCTATTCTGCTGATGATATCAACAAATTTCTCT 2029
DB 1029 TTTTGGTATCTTTGGCGGTGTGATCGCTATTCTGGCGATGATATCAACAGTTCCCTGT 1088
QY 2030 GGATGTTAGATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGACTATTTTA 2089
DB 1089 GGATGTTCCGAATTTGCTGAGGAGAGATCCCAAGGACATTAAGGAAAGCGATTAATT 1148
QY 2090 CCCACAGGAGAGATTCGGGTGACAAAGCAGATCCCTACTTTTGTGAAATTCGCTTA 2149
DB 1149 CCGACCGCGGTGAATTCAGGATAGTCCGAGAGTGTCCGCGCTGTCTCACTGCTTA 1208
QY 2150 TGTATAAATGCTACTACAGATTTGGAGAAATGAGCTGGATTTTCGTACACCCGAG 2209
DB 1209 TGTACAAATTAAGCTACTACAGATTCGGGGAATTTGAATTTGGACTACAGAGTCCATCTG 1268
QY 2210 GTTTTGACCGAACACGTAAATGCTGAGATTTGGAATTAAGGACATTAATTAACAACTTTGG 2269
DB 1269 GATATGCTGACACGTAAACGCGCTCATTTGGGATTAAGACTTCGATCTGACCTACCTGG 1328
QY 2270 AAGAGGCTTTTACATGACACATCTGGCTTTTATAGGATATATAAGTAAAGCACTGTATA 2329
DB 1329 AGGAGGCTTACACCAAGACATCTGGCTTTTCTGCTATCTATAGGTTGAAGAGCGCATG 1388
QY 2330 ACAGGAGACATTTAGATCACAACCTTCGAGTCCACCAACATTTTCCCAAAACAGAGATT 2389
DB 1389 AGTTCAATAGACCTACTAGACACCAAGAGAGAGAGATTCCTCCAGCAACTCAATTT 1448
QY 2390 TGTCAAGAGACTTACCAAGAGAGGAGGCTGGCTGCTACATTAATAAGCTGGTTTAAAGA 2449
DB 1449 CGAGAGAGAACTTAAAGCGTCGCAAG---GGCTACATACGAAACCGACCGGTTGTTTA 1505
QY 2450 AAGGCAAGAAATCTAAGAGAGAC 2474
DB 1506 AGGNAACGAACTTGAATTAAC 1530

```

```

RESULT 7
US-09-270-767-11648
; Sequence 11648, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11648

```

```

; LENGTH: 1386
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-11648

Query Match      12.5%; Score 309.6; DB 4; Length 1386;
Best Local Similarity 56.9%; Pred. No. 3.5e-73;
Matches 567; Conservative 0; Mismatches 429; Indels 0; Gaps 0;

QY 129 GCGCGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 188
DB 386 GGCAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 445
QY 189 GCAGCGCGGCTGGGTGGCAGCTCGCTTCTCTCTTCCATCCATCCATCCATCCATCCAT 248
DB 446 CACCTGGGACAAACAGAGAGACCTGCTCAAGCTGGCCATCTCTCATCTCTGGCAGCGG 505
QY 249 CGGCTTCAGCTCGCGCTCTTCGCGCTCATCCGCTTCGAAAGCATCATCCACGAGTTTGA 308
DB 506 ATCAATTTGCACACAGCTTGTCTCTGTCTGGATTCGAAAGCATCATCCATGAGTTTGA 565
QY 309 CCCTGGTTTAACTATPAGATCAACATCATCTTGCATCTCATGGGTTTCTAATGAAATTT 368
DB 566 TCCGTACTTCAACTACCGCACCGCGGTTCTGCGGAGCAGGCGCTTTTACAAGTTTCCA 625
QY 369 AATTTGGTTGATGAAGAGCATGTTATCCACTAGAGAGATAGTAGGTGCTACTGTTTA 428
DB 626 CAATCTGTTGATGACCGCGCTGTTATCCCTGGCGCATCATCGCGCGCACATCTA 685
QY 429 CCAGAGGTTGATGATAACCGCTGGCTTTATTCATTTGGAATTTTAAATACATTAACA 488
DB 686 TCCCGGCTGATGCTCACCTCGCGCGCTGTACCGCTGATGTGGCTGCTCAATGTGAC 745
QY 489 TGTTCATATAAGAGAGATGTTGCTCTCTGCGACCACTTTTAGCGGCTTACATCTAT 548
DB 746 CATCGACATACGAGAGCTGCTGCTCTCTGCGCGCTTCTCTCTCTGCTGACACGCT 805
QY 549 ATCTACTTCTCTGCTTACAGAGAACTTTTGAACCAAGAGCAGGACTTTTAGCTGTTG 608
DB 806 GGTGACCTACGCGCTCACAAAGAGATACACAGCACTGGAGCTGGCTGGCGCGCGCG 865
QY 609 TTTTATGCTATGTTACAGAGCTACATATCTCGGTCAGTCTGATCTCTTATGATTAATGA 668
DB 866 TTTGATATCATGTTCCCGGTTATCTCTCGATCCGTTGGGAGATCGTACGACATGA 925
QY 669 AGCATTTGCTATTTTTCACCTTCAGTTTCACTATTTATTTGGGTAAATCTGTAAATAAC 728
DB 926 AGCATTCGCTATTTTTCGATGCTCTTCACTTACTATTTTGGATCAAGCGCGTAAAGAC 985
QY 729 TGGGTCAGTTTTCGACAACTGCTGCTGCTTATCTCTTATCTATATGCTCTGCTTG 788
DB 986 GGGCAGATCTTTTGTGCGGCTATGTCGCGATTTGCGCTTCTATATGTTCTCTCTG 1045
QY 789 GGGTGGTTATGTTTATCATCAATCTTATTCCTGCTGATGTTTGTGTTTACTGAT 848
DB 1046 GGGTGGCTATGCTTCTCTGATTAACCTAATCCGCTGCGACGCTGCTGGGCTGATGATC 1105
QY 849 GCAGAGTACGCAAGAGAGCTTACATAGCATATAGCACTTTTCTACATTTGGGTTTAA 908
DB 1106 CGGACGTTTCTGCGACAGGATCTATAGCATACAGAGCTTACTTCTGCTGCGACCAT 1165
QY 909 ATTATCAATGCGAGATACCTTTTGTGGATTCAGGCAATACAGAACTGAACATGCGC 968
DB 1166 TCTGTGATGCGAGATCTCGTTTGTGGGATTCGAAACCATCCAGAGCTCCGAAACATG 1225
QY 969 AGCTGCGAGTCTTTTTCGATTCCTCAAGCTTATCTTCTTCTGAGTATCTGAGAGCCG 1028
DB 1226 GGCACGAGAACTTTTGGCTGTGCGATTCGAGATTCAGCTTTTCTGCTGCTGCTG 1285
QY 1029 ATTAAACAAACAGAGTTCCAGACCTTTTCTTTTGGGTGATACATGCTGAGGTG 1088
DB 1286 CATTCGCAAGGATCACTTCGATCTGCTTCAAGAGCTTGGTTTCCAGTGTTTTACTGT 1345

```


SEQ ID NO 1965
LENGTH: 507
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 122..505
US-09-513-999C-1965

Query Match 20.4%; Score 507; DB 4; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.2e-126;
Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1683 CTACTCTAGTCCCAAGTCTAGTCTGGCTCATACATCATGATGGCACCAGGAATATCTT 1742
DB 1 CTACTCTAGTCCCAAGTCTAGTCTGGCTCATACATCATGATGGCACCAGGAATATCTT 60
QY 1743 AGATGATTTTAGAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGCACGAGT 1802
DB 61 AGATGATTTTAGAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGCACGAGT 120
QY 1803 AATGCTTGGTGGGATTTAGCTATCAGATAGCTGGAATGGCTTAATAGAACTACGTTGGT 1862
DB 121 AATGCTTGGTGGGATTTAGCTATCAGATAGCTGGAATGGCTTAATAGAACTACGTTGGT 180
QY 1863 GGATAATAACACCTGGGAATTAACAGCCACATAGCACCTGGTGGGAAAGCTATCTTCTAA 1922
DB 181 GGATAATAACACCTGGGAATTAACAGCCACATAGCACCTGGTGGGAAAGCTATCTTCTAA 240
QY 1923 TGAACAGCAGCCTTAAATATCATGAGGCTCTAGATGATGATTTTGGTTATTTT 1982
DB 241 TGAACAGCAGCCTTAAATATCATGAGGCTCTAGATGATGATTTTGGTTATTTT 300
QY 1983 TGGAGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTGGATGGTTAGAT 2042
DB 301 TGGAGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTGGATGGTTAGAT 360
QY 2043 AGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGGAGA 2102
DB 361 AGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGGAGA 420
QY 2103 ATTCCGTTGTAGAAAAGCAGGATCCCGCTACTTTTGGATTTGGCTTATGATATAAAATGTC 2162
DB 421 ATTCCGTTGTAGAAAAGCAGGATCCCGCTACTTTTGGATTTGGCTTATGATATAAAATGTC 480
QY 2163 ATACTACAGATTGGGAAATGAGCT 2189
DB 481 ATACTACAGATTGGGAAATGAGCT 507

RESULT 3

US-09-328-111-332
Sequence 332, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Derti, Adrian
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CDD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801

EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 332
LENGTH: 616
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(616)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-332

Query Match 16.7%; Score 415.4; DB 3; Length 616;
Best Local Similarity 95.8%; Pred. No. 6.5e-102;
Matches 436; Conservative 0; Mismatches 18; Indels 1; Gaps 1;
QY 622 GTACCAAGGCTACATATCTCGGTACGTAGCTGGATCCCTTTGATATGAAGGCAATGCTATT 681
DB 5 GTACCAAGGCTACATATCTCGGTACGTAGCTGGATCCCTTTGATATGAAGGCAATGCTATT 64
QY 682 TTTCACCTTCAGTTCACATACCTATTTATGGGTAAATCTGTAAAACTGGGTCAAGTTTTT 741
DB 65 TTTCACCTTCAGTTCACATACCTATTTATGGGTAAATCTGTAAAACTGGGTCAAGTTTTT 124
QY 742 TGGACAATGTGCTGCTTATCTTATCTTATATGCTCTGCTGGGGTGGTTATGTA 801
DB 125 TGGACAATGTGCTGCTTATCTTATCTTATATGCTCTGCTGGGGTGGTTATGTA 184
QY 802 TTTCATCATCAATCTTATTCCTACCTGCTATTTTGTGTGTACTGATGCAGAGATACAGC 861
DB 185 TTTCATCATCAATCTTATTCCTACCTGCTATTTTGTGTGTACTGATGCAGAGATACAGC 244
QY 862 AAAAGAGTCTACATGACATATAGCACTTTTACATTTGGGTTTAAATTTATCAATGCAG 921
DB 245 AAAAGAGTCTACATGACATATAGCACTTTTACATTTGGGTTTAAATTTATCAATGCAG 304
QY 922 ATACCTTTTGTGGGATTCAGCAATCAAGCAAGTGAACATGCGAGCTGCAGGTGC 981
DB 305 ATACCTTTTGTGGGATTCAGCAATCAAGCAAGTGAACATGCGAGCTGCAGGTGC 364
QY 982 TTTCGATTTCTGCAAGCTTATGCTTTTTCGAGTATCTGAGAGACCGATTAACAAACAA 1041
DB 365 TTTCGATTTCTGCAAGCTTAACTTTCTTGCAGTATCTGAGA-ACCGATTACCAACAA 423
QY 1042 GAGTTCACACCTTTTCTTTTGGGTGTATCACT 1076
DB 424 GAGTTCACACCTTTTCTTTTGGGATCTACT 458

RESULT 4

US-09-270-767-12331
Sequence 12331, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12331
LENGTH: 900
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-12331

Query Match 14.6%; Score 361.8; DB 4; Length 900;
Best Local Similarity 66.5%; Pred. No. 2.3e-87;
Matches 519; Conservative 0; Mismatches 262; Indels 0; Gaps 0;
QY 191 AGCCGGCTGGTGGCAGTCGCTTCTCTCTTACCACCTCTCTCTCTGGCTGGCTTCCG 250

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 00:40:24 ; Search time 198.055 Seconds
(without alignments)
8903.949 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagccctcgccccc.....tatctaagaagactgttttaa 2481

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 35539441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	539.4	21.7	2157	4	US-09-614-221A-318
2	507	20.4	507	4	US-09-513-999C-1965
3	415.4	16.7	616	3	US-09-328-111-332
4	361.8	14.6	900	4	US-09-270-767-12331
5	340.2	13.7	1660	4	US-09-270-767-158
6	340.2	13.7	1660	4	US-09-270-767-15440
7	309.6	12.5	1386	4	US-09-270-767-11648
8	273.8	11.0	1660	4	US-09-270-767-15440
9	273.8	11.0	1660	4	US-09-270-767-15440
10	272.2	11.0	867	4	US-09-248-796A-3089
11	250	10.1	250	4	US-09-513-999C-1438
12	242.4	9.8	245	4	US-09-513-999C-21090
13	237.4	9.6	914	4	US-09-270-767-12856
14	201.4	8.1	560	4	US-09-270-767-1318
15	201.4	8.1	560	4	US-09-270-767-16600
16	176.6	7.1	594	4	US-09-248-796A-3090
17	173.8	7.0	503	4	US-09-270-767-27262
18	149.4	6.0	487	4	US-09-702-705-1655
19	149.4	6.0	487	4	US-09-736-457-1655
20	149.4	6.0	487	4	US-09-614-124B-1655
21	149.4	6.0	487	4	US-09-671-325-1655
22	149.4	6.0	487	4	US-09-658-824-1655
23	147	5.9	307	4	US-09-702-705-375
24	147	5.9	307	4	US-09-702-705-1271
25	147	5.9	307	4	US-09-736-457-375
26	147	5.9	307	4	US-09-736-457-1271
27	147	5.9	307	4	US-09-614-124B-375

Sequence 1271, App
Sequence 375, App
Sequence 1271, App
Sequence 375, App
Sequence 375, App
Sequence 1271, App
Sequence 588, App
Sequence 588, App
Sequence 588, App
Sequence 588, App
Sequence 588, App
Sequence 213, App
Sequence 4442, App
Sequence 4834, App
Sequence 6705, App
Sequence 6869, App
Sequence 271, App

28 147 5.9 307 4 US-09-614-124B-1271
29 147 5.9 307 4 US-09-671-325-375
30 147 5.9 307 4 US-09-671-325-1271
31 147 5.9 307 4 US-09-589-184-375
32 147 5.9 307 4 US-09-658-824-375
33 147 5.9 307 4 US-09-658-824-1271
34 145.4 5.9 307 4 US-09-702-705-588
35 145.4 5.9 307 4 US-09-736-457-588
36 145.4 5.9 307 4 US-09-614-124B-588
37 145.4 5.9 307 4 US-09-671-325-588
38 145.4 5.9 307 4 US-09-589-184-588
39 145.4 5.9 307 4 US-09-658-824-588
40 137.4 5.5 487 3 US-09-385-982-213
41 134 5.4 292 4 US-09-313-294A-4442
42 123 5.0 291 4 US-09-313-294A-4834
43 104.6 4.2 288 4 US-09-313-294A-6705
44 100.8 4.1 302 4 US-09-313-294A-6869
45 96 3.9 414 4 US-09-513-999C-271

ALIGNMENTS

RESULT 1

US-09-614-221A-318
; Sequence 318, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunanandaa, Balasulojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614, 221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 318
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-318

Query Match	21.7%	Score 539.4;	DB 4;	Length 2157;
Best Local Similarity	55.1%	Pred No. 5.3e-135;		
Matches 1237;	Conservative	0;	Mismatches 896;	Indels 110; Gaps 5;
Qy	175	TCGGGGGGCTCTCGACGCCGCTGGTGGGAGTGGTCTCTCTCTCACCATCTCTTC	234	
Db	7	TCGACCGGCTGCTGTTTCTGTTTTCAGACCATCTCTCAAGCTCGTCTCTG	66	
Qy	235	CTGGCTGGCTTGGCGGCTTCAGCTCGGCTCTTTCGGCGTCTCTCGAAGGATC	294	
Db	67	GGATTTTGGGGCTGCCATATCATCAGTTTGTTCAGTCAATTTGAGTCTATT	126	
Qy	295	ATCCAGAGTTCGACCCGCTGGTTAACTATAGATCAACATCATCTTGATCTCGG	354	
Db	127	ATCCATGAATTCGACCCCTCTGTTCAATATATAGGCTACCAATATCTCGTCAACAATTC	186	
Qy	355	TTCTATGAATTTTAAATTTGTTGATCAAAAGAGCATGTATCCACTAGGAAGATAGTA	414	
Db	187	TTTACAGTTTGTGAACTGTTTGACGACCGTACCTGTACCCCTCGNAGGTTACT	246	
Qy	415	GGTGTACTGTTTACCCAGGTTGATGATAACCCGCTGGCTTATT---CATGTGATTTTA	471	
Db	247	GGAGGACTTTTATATCTGTTTGTATGACGACTAGTGGTTCATCTGGCAGCCCTCGC	306	
Qy	472	AATACATTCAACATAAATCTTACATAGACAGCTATGTGTGTTTCCCTTGACCAACTTTT	531	
Db	307	AACTGGTGGGTTGCCCATTTGACATCAGAACGTTTGTGTGCTATTGCGCACTATT	366	

HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017376-A/29
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253173
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, TOMOYASU
 SUGIYAMA,
 PI KOJI HAYASHI
 FC
 C12N15/09, C07K14/47, C07K14/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
 10,
 PC C12P21/02, C12Q1/68, C12P21/08, C12N15/00, C12N5/00 CC
 Secretory protein or membrane protein
 FH Key Location/Qualifiers
 FT CDS (94)..(951).
 FEATURES
 source
 1..2510
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Query Match 38.5%; Score 954; DB 6; Length 2510;
 Best Local Similarity 100.0%; Pred. No. 2e-178;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1528 TTGTATGATAAGGAGGAGTAAAGTGAAGGAAACATGCAACATGAACAGAGAAAAAAGTGAAGAG 1587
 DB 1 TTGTATGATAAGGAGGAGTAAAGTGAAGGAAACATGCAACATGAACAGAGAAAAAAGTGAAGAG 60
 QY 1588 GGATTAGCCCTATATATAAAGGAGTGTCCACATGTTGATGCTGATGCTATGATGATG 1647
 DB 61 GGATTAGCCCTATATATAAAGGAGTGTCCACATGTTGATGCTGATGCTATGATGATG 120
 QY 1648 TTGCTGTCCACTGTACCTGGTCAAGCAATGCCCTACTCTAGTCCAAAGTGTAGTCTG 1707
 DB 121 TTGCTGTCCACTGTACCTGGTCAAGCAATGCCCTACTCTAGTCCAAAGTGTAGTCTG 180
 QY 1708 GCCTCATACATCATGATGACAGGAGATATCTTAGATGATTTTAGAGAGCTTACTTT 1767
 DB 181 GCCTCATACATCATGATGACAGGAGATATCTTAGATGATTTTAGAGAGCTTACTTT 240
 QY 1768 TGGCTAAGGCAAAATACAGATGAACATGACAGAGTAATGCTTTGGTGGGATTATGGCTAT 1827
 DB 241 TGGCTAAGGCAAAATACAGATGAACATGACAGAGTAATGCTTTGGTGGGATTATGGCTAT 300
 QY 1828 CAGATAGCTGGATGGCTTAATAGACTACGTTGGTGATATACACCTGGATTAACAGC 1887
 DB 301 CAGATAGCTGGATGGCTTAATAGACTACGTTGGTGATATACACCTGGATTAACAGC 360
 QY 1888 CACATAGCACTGGTGGGAAAAGTATCTTCTTAATGAACACAGCAGCCTATAAAATCATG 1947
 DB 361 CACATAGCACTGGTGGGAAAAGTATCTTCTTAATGAACACAGCAGCCTATAAAATCATG 420
 QY 1948 AGGACTCTAGATGATGATATGTTTGGTATTTTGGAGGGTATTTGGCTATTTCTGGT 2007
 DB 421 AGGACTCTAGATGATGATATGTTTGGTATTTTGGAGGGTATTTGGCTATTTCTGGT 480
 QY 2008 GATGATATCAACAAATTTCTGATGGTTAGGATAGCTGAAGAGAAACATCCCAAGAC 2067
 DB 481 GATGATATCAACAAATTTCTGATGGTTAGGATAGCTGAAGAGAAACATCCCAAGAC 540
 QY 2068 ATTGGGAAAGTACTATTTTACCCACAGGAGATTTCCGTGTAGACAAAGCAGATCC 2127
 DB 541 ATTGGGAAAGTACTATTTTACCCACAGGAGATTTCCGTGTAGACAAAGCAGATCC 600
 QY 2128 CCTACTTTGTTGAATGCTTATGATATAAATGTCATACAGATTTGGAGAAATGCAG 2187
 DB 601 CCTACTTTGTTGAATGCTTATGATATAAATGTCATACAGATTTGGAGAAATGCAG 660
 QY 2188 CTGGATTTTGTACACCCCAAGGTTTGGACGACAGTAACTGCTGAGATTTGGAATAG 2247
 DB 661 CTGGATTTTGTACACCCCAAGGTTTGGACGACAGTAACTGCTGAGATTTGGAATAG 720

2248 GACATTAAATTCAAACATTTTGAAGAAGCCTTTACATCAGAAACACTGGCTTTTAGGATA 2307
 DB 721 GACATTAAATTCAAACATTTTGAAGAAGCCTTTACATCAGAAACACTGGCTTTTAGGATA 780
 QY 2308 TATAAAGTAAAGACACCTTGATAAACAGGAGACATTAGATCAAAACCTCGAGTCACCAAC 2367
 DB 781 TATAAAGTAAAGACACCTTGATAAACAGGAGACATTAGATCAAAACCTCGAGTCACCAAC 840
 QY 2368 ATTTTCCCAAAACAGAGATTTTGTCAAGAAGACTACCAAAAGAGAGCGTGGCTACATT 2427
 DB 841 ATTTTCCCAAAACAGAGATTTTGTCAAGAAGACTACCAAAAGAGAGCGTGGCTACATT 900
 QY 2428 AAAAATAAGCTGGTTTTTAAAGAAAGCAAGAAATATCTTAAGAAGACTGTTTTAA 2481
 DB 901 AAAAATAAGCTGGTTTTTAAAGAAAGCAAGAAATATCTTAAGAAGACTGTTTTAA 954

Search completed: December 15, 2004, 07:01:19
 Job time : 10331.1 secs

QY 2156 AAATGTCATACAGATTTGGAGAAATGCGAGTGGATTTTCGTACACCCCGAGTTTG 2215
 Db 2091 AATTAAGCTACTACAGATTCGGGAATTTGAATGGAGTCCATCTGGATATG 2150
 QY 2216 ACCGACACGTAATCTGAGATTTGGAATAGGACATTAATCAACATTTTGGAGAG 2275
 Db 2151 ATCGCACACGTAACCGCTCATTTGGGAATAGGATTCGATCTGACCTACCTGGAGAG 2210
 QY 2276 CTTTACATCAGAACACATCTGGTGTGTTAGGATATATAAAGTAAAGCACCTGTATAACAGGG 2335
 Db 2211 CCTACACACAGAACATCTGGTGTGTTAGGATATATAAAGTAAAGCACCTGTATAACAGGG 2335
 QY 2336 AGACATTTAGTACACAAACCTCGAGTACCAACATTTTCCCAAAACAGAGATTTTGTCAA 2395
 Db 2271 ATAGACCATCTAGACACCAAGAGAGAGAGATTCCTCCAGCAACTTCATTTCCAGAA 2330
 QY 2396 AGAAGACTACCAAAAGGAGCGTGGTACATATAAATAAGCTGGTGTATTAAGAAAGGCA 2455
 Db 2331 AGAAGCTTAAGCGTGGCAAG---GGCTACATACGAAACCGACCGGTTGTTGTTAAGGAA 2387
 QY 2456 AGAAATATCTAAGAGAC 2474
 Db 2388 AACGAACCTTGAATTAAC 2406

RESULT 14

AX136135 2510 bp DNA linear PAT 30-MAY-2001
 LOCUS
 DEFINITION Sequence 57 from Patent EP1067182.
 ACCESSION AX136135
 VERSION AX136135.1 GI:14272543

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Ota.T., Isogai.T., Nishikawa.T., Kawai.Y., Sugiyama.T. and

Hayashi.K.

Secretory protein or membrane protein

Patent: EP 1067182-A 57 10-JAN-2001;

Helix Research Institute (JP)

Location/Qualifiers

1..2510

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

94..954

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC39710.1"

/translation="MLMLMLMFAVHCTWVTSVSVSLVSYNHDGTRNLLDDF

REAYFLQRNTDEHARVMSWDYGYQAGVANRTTLVDNNWNHSHIALVGMKSNSE

TAAYKIMRTPLDVYVLVFGVIGYSGDINKFLMVRVIAEGHPKIDIESDYFTPOG

EPKVDKAGSLDNLCLMYKMSYRFGEMQLDFRTPGDFRTRNAEIGNDKIDFKHLEE

ARTSEHLVRLVYKVKAPDNRETLDKPRVTNIPPKQKYLKSKTKRGVINKKLVEK

KGKSKSKTV"

ORIGIN

Query Match 38.5%; Score 954; DB 6; Length 2510;
 Best Local Similarity 100.0%; Pred. No. 2e-178;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1528 TTGTATGATAGGACAGGTAAAGTGGAGAAACATGCACTAACAGGAAAAAACTGAAGAG 1587

Db 1 TTGTATGATAGGACAGGTAAAGTGGAGAAACATGCACTAACAGGAAAAAACTGAAGAG 60

QY 1588 GATTTAGGCCCTTAATATAAAGAGATTTGTACCCATGTTGATGCTGATGCTATTGATGATG 1647

Db 61 GGATTTAGGCCCTTAATATAAAGAGATTTGTACCCATGTTGATGCTGATGCTATTGATGATG 120

QY 1648 TTGCTGTCCACTGTACTCTGGGTCAAGAGAAATGCTACTCTAGTCCAAGTGTAGTCTG 1707
 Db 121 TTGCTGTCCACTGTACTCTGGGTCAAGAGAAATGCTACTCTAGTCCAAGTGTAGTCTG 180
 QY 1708 GCCTCATACAATCATGATGGCACACAGGAATATCTTAGATGATTTTAGAAGAGCTTACTTT 1767
 Db 181 GCCTCATACAATCATGATGGCACACAGGAATATCTTAGATGATTTTAGAAGAGCTTACTTT 240
 QY 1768 TGCTTAAGGCAAAATACAGATGAATGACAGAGTAATGCTCTGGTGGGATTTAGGCTAT 1827
 Db 241 TGCTTAAGGCAAAATACAGATGAATGACAGAGTAATGCTCTGGTGGGATTTAGGCTAT 300
 QY 1828 CAGATAGCTGAATGGCTAATAGAACTACTGTTGGTGGATAATAACACCTGGAATAACAGC 1887
 Db 301 CAGATAGCTGAATGGCTAATAGAACTACTGTTGGTGGATAATAACACCTGGAATAACAGC 360
 QY 1888 CAGATAGCTGGTGGGAAAGCTATGCTCTTATGAAACAGCAGAGCTATAAATCATG 1947
 Db 361 CAGATAGCTGGTGGGAAAGCTATGCTCTTATGAAACAGCAGAGCTATAAATCATG 420
 QY 1948 AGGACTCTAGATGATGATTTTGGTGTATTTTGGAGGGGTTATTTGGCTATTTCTGGT 2007
 Db 421 AGGACTCTAGATGATGATTTTGGTGTATTTTGGAGGGGTTATTTGGCTATTTCTGGT 480
 QY 2008 GATGATATCAACAAATTTCTCTGGATGTTAGTAGTGAAGGAGAACATCCCAAAGAC 2067
 Db 481 GATGATATCAACAAATTTCTCTGGATGTTAGTAGTGAAGGAGAACATCCCAAAGAC 540
 QY 2068 ATTCCGGAAAGTACTATTTTACCCACAGGAGAAATTCCTGTAGACAAAGCAGGATCC 2127
 Db 541 ATTCCGGAAAGTACTATTTTACCCACAGGAGAAATTCCTGTAGACAAAGCAGGATCC 600
 QY 2128 CCTACTTTGTTGAATGCTTATGATATAAATGTCATACTACAGATTTGGAGAAATGCAG 2187
 Db 601 CCTACTTTGTTGAATGCTTATGATATAAATGTCATACTACAGATTTGGAGAAATGCAG 660
 QY 2188 CTGATTTGCTACACCCCGAGTTTGAACGAAACAGTAACTGAGATTTGGAATAAG 2247
 Db 661 CTGATTTGCTACACCCCGAGTTTGAACGAAACAGTAACTGAGATTTGGAATAAG 720
 QY 2248 GACATTAATTCAAACATTTGGAAGAGCTTTTACATCAGAACACTGCTTGTAGGATA 2307
 Db 721 GACATTAATTCAAACATTTGGAAGAGCTTTTACATCAGAACACTGCTTGTAGGATA 780
 QY 2308 TATAAGTAAAGCACCTGATACAGGAGAGATTTAGATCAACAACTCGAGTCAACAC 2367
 Db 781 TATAAGTAAAGCACCTGATACAGGAGAGATTTAGATCAACAACTCGAGTCAACAC 840
 QY 2368 ATTTTCCCAAAACAGAAAGTATTTGTCAAAGAGAGCTACCAAAAGGAGAGCTGGCTACATT 2427
 Db 841 ATTTTCCCAAAACAGAAAGTATTTGTCAAAGAGAGCTACCAAAAGGAGAGCTGGCTACATT 900
 QY 2428 AAAAAAGCTGCTTTTAAAGAACCAAGAAATATCTTAGAAGACTGTTTAA 2481
 Db 901 AAAAAAGCTGCTTTTAAAGAACCAAGAAATATCTTAGAAGACTGTTTAA 954

RESULT 15

BD123520

LOCUS

BD123520

DEFINITION

Secretory protein or membrane protein.

ACCESSION

BD123520.1

VERSION

BD123520.1

KEYWORDS

JP 2002017376-A/29.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Ota.T., Isogai.T., Nishikawa.T., Kawai.Y., Sugiyama.T. and

Hayashi.K.

Secretory protein or membrane protein

Patent: Jp 2002017376-A 29 22-JAN-2002;

JOURNAL

TITLE Detection kits, such as nucleic acid arrays, for detecting the expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL Patent: WO 0171042-A 17111 27-SEP-2001;
PE Corporation (NY) (US)

FEATURES
source 1..2699
Location/Qualifiers
/organism="Drosophila sp."
/mol_type="unassigned DNA"
/db_xref="taxon:7242"

ORIGIN
Query Match 38.5%; Score 955.8; DB 6; Length 2699;
Best Local Similarity 64.8%; Pred. No. 8.8e-179;
Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;

Qy 191 AGCGGCTGGGTCAGTCGCTTCCTTCACCAATCTCTTCCTGGCTGGCTGGCG 250
Db |||||
Qy 111 AGTGGCTGGCTACAGCAGCCTAATCACTTCGCCATCTGCTAATPCGCTGGCGG 170
Db |||||
Qy 251 GCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAAAGCATCACTCCAGATTGACC 310
Db |||||
Qy 171 GATTTCTCTCGCTCTTCGCGCTCATCGCTTCGAGTCGATTCATGATGATGATC 230
Db |||||
Qy 311 CGTGGTTAACTATAGATCAACACATCATCTTCATCTCATGGCTTCATGATTTTAA 370
Db |||||
Qy 231 CGTGGTTCAACTACCGGCGCCACCGCTTACATGCTGCGAGATGGTTGGTAACTTCTCA 290
Db |||||
Qy 371 ATTGGTTTGATGAAGAGCATGGTATCCACTAGAGAGATAGTAGTGGTACTGTTACC 430
Db |||||
Qy 291 ACTGGTTCAGAGCGCGCATGATATCCGCTCGCAGGATGTTGGCGGTACCGTCTATC 350
Db |||||
Qy 431 CAGGTTGATGATCAACCGCTGGCTTATCATTTGATTTTAAATACATGAAACATATG 490
Db |||||
Qy 351 CGCGCTGATGATTAACGTCGCGCGGAATCAATGGCTGCTGCACTCAACATACCGG 410
Db |||||
Qy 491 TTCATGAAGAGCATGATGTTCTTCGCAACAACTTTAGCGGCTTACATCTATAT 550
Db |||||
Qy 411 TCCATATTCGTAATCATCTGGTGTCTCGCGCGGATCTTCAGTGGCTGACCTCCATCT 470
Db |||||
Qy 551 CTACTTCTCTCTTACAGAGAACTTTGGAAACAGAGCAGGACATTTTATGCTGTTGT 610
Db |||||
Qy 471 CCACCTACTCTCTGACCAAGAGCTGTGGTCCGCGGCGCGCGCTTTCGCGCGAGCT 530
Db |||||
Qy 611 TTATTTGATTTGACAGGCTACATATCTCGCTCAGTATGCTGATCTTGTATATGAAG 670
Db |||||
Qy 531 TCATCGCATCTGCTGCTGCTGCTACATCACTAGTGGCTGGCTGGATGCTACAGATACGAGG 590
Db |||||
Qy 671 GCATGCTATTTTGGCATCTCAGTTCATCACTATTTATGGGTAATCTGTAAACATG 730
Db |||||
Qy 591 GCATTTGCCATATTCGCGCTGCTGCTCACTTCTCTGCTGGCTGCTCAGTGAAGACTG 650
Db |||||
Qy 731 GGTGATTTTGGCAATGCTGCTGCTTATCTCTTATATGCTCTCTGCTGGG 790
Db |||||
Qy 651 GATCGGTCTCTGCTGCGCGCAGCGCTTGTCTTCTTCTTCTTCTGCTGGG 710
Db |||||
Qy 791 GTGGTATGATTTATCATCAATCTTATTCATCTGATGATTTTGTGTTTACTGATGC 850
Db |||||
Qy 711 GTGGTATGATTTATCATCAATCTTATTCATCTGATGATTTTGTGCTCTTCTTCTG 770
Db |||||
Qy 851 AGAGATACAGCAAAAGAGTCTACATGATATAGCACTTCTTCACTTGTGGGTTTAAAT 910
Db |||||
Qy 771 GCAGGTACTCGCGGCTCTGCTGACAGCTACAGACCTTCTTACCTCTGGGACTGCT 830
Db |||||
Qy 911 TATCAATCAGATACCTTTTGTGGGATTCAGCGCAATCAGAAAGTGAACATGGGAG 970
Db |||||
Qy 831 TCTCCATGAGATCCCTTCTGTTGGGATTCGAACCGGATACGACAGTGAACATGGCTG 890
Db |||||
Qy 971 CTGAGGCTCTTTCGATTTGCTGCAAGCTTATGCTTCTTTCGAGTATCTGAGAGCCGAT 1030
Db |||||
Qy 891 CGCTGGAGTGTGTTGCTCTTATGCGCGTGGCGACCTTTCGCGCAATTTGCACTCGGTCG 950
Db |||||
Qy 1031 TAACAAAACAGAGTTCCAGACCCCTTTCTTTTGGGTGATACATAGCTGCGAGTGTG 1090
Db |||||

Db 951 TGTGCGCAACGAGTTCCGAGAGCTGTTTCATCTGCTGGCGGATTCCTGCTGGCGCTGGCG 1010
Qy 1091 TGTTCCTTATGCTCATCTATTTTACTTATACAGGTTTACATTTGCACCATGAGTGGCAGGT 1150
Db 1011 TCTTTGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1070
Qy 1151 TTTATTCATTGTTGGGATGCTGGTATGCAAAATATACATTCATTCATTCATTCATTCAT 1210
Db 1071 TCTACTCGCTGTTGGGATGCTGGTATGCAAAATATACATTCATTCATTCATTCATTCAT 1130
Qy 1211 CTGAGCATCAACTACGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1270
Db 1131 CGGAGCATCAGCCACCACTTGGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1190
Qy 1271 CTTTCCACGAGGCTTGGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1330
Db 1191 CTTTCCACGAGGATGTTGGTATGCAAAATATACATTCATTCATTCATTCATTCATTCAT 1250
Qy 1331 CTTATATCAATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1390
Db 1251 TGTCTGACCCATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1310
Qy 1391 TGAATCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1450
Db 1311 TCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1370
Qy 1451 TGGGGGATG-----CATGAAAGGAGAAATCCACTGTGGAG-----GACAGCAGTG 1498
Db 1371 TGCAAGAGGATTCGCTTAAGCGATGGGACACAGGACATAGCGCAGGACCAAGTGGATG 1430
Qy 1499 ATGAGGATGACAAAGCAACCAAGGAAATTTGATGATGAAAGGAGGTAAGTGAAGGAA 1558
Db 1431 AAGCTGAGGATTCATTTGAGAAAGAGAGCTGACGACAGGCTGGCAAGCTGAGCATC 1490
Qy 1559 ATGCAACTCAACAGGAAAGAACTGAAGAGGATTTAGGCCCTTAAATATATAAGCAATTC 1618
Db 1491 GTACTAAGCATGATGCCAGCAGGATCTGCGCTCAGCTCCACCTGAAGATTTGTTA 1550
Qy 1619 CCATGTTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1678
Db 1551 TTTTGGCGCTTCTAATGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 1610
Qy 1679 ATGCTTACTTACTGCTCAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1735
Db 1611 ATGCTTACTTACTGCTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1670
Qy 1736 ATATCTTATGATGATTTAGAGAGCTTACTTTTGGCTTAAAGCAAAATACAGATGAATG 1795
Db 1671 ACATTTTATGAGGATTTTACAGAGGCTTACTTCTGCTTTCGCAAGCACTGCGCATG 1730
Qy 1796 CAGGATTAATGCTTGGTGGGATTTATGCTATCAGATAGCTGGAATGGCTAATAGAACTA 1855
Db 1731 CTGCGCTTATGCTTGGTGGGATTTACGATACAGATAGCGGAAATGCAAAACAGAACTG 1790
Qy 1856 CGTTGGTGAATTAATACACCTTGAATTAACAGCAATAGCAGTCTGCTGGGAAAGCTATGT 1915
Db 1791 CGTATGATGATTAATTAATGAGGAAATAGTCAATAGCCTGCTGCTGGCAAGCAATGT 1850
Qy 1916 CTTTCTAATGAAACAGCAGCCTTAATAAATCATGAGGACTCTAGATGTAGATTAATTTTGG 1975
Db 1851 CTTTCAACCGAGGAGAGTCTTACGAAATTTATGACATCTTCTGAGCTGAGCTTCTTGG 1910
Qy 1976 TTTATTTTGGAGGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2035
Db 1911 TGATCTTTGGCGGTGATCGGCTATTTCTGGGATGATATCAACAGTTCTCTGTTGATGG 1970
Qy 2036 TTAGGATAGCTGAGGAGAACTCCCAAAGCATTTCCGGAAGTGAATTTTATTTTCCACC 2095
Db 1971 TCCGAATTTCTGAGGAGAGCATCCCAAAGCATTTAAGGAAAGGCAATTTACTTTTCCACC 2030
Qy 2096 AGGGAATTTCTGCTGAGCAAGAGGATCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2155
Db 2031 CGGCTGAATTTACGGGTGATGTCGCAAGGCTGCTCCGGGCTTCTCTCACTGCTCTTATG 2090

Db 462 CCACCTACTCTGACCAAGAGCTGGTCCGGGCGCGCCCTCTTGGCGCCAGCT 521
Qy 611 TTATTGCTATTGTACACGCTACATATCTCGGTGAGTGGATCCTTTGATTAAGAG 670
Db 522 TCATCGCATCGTGGCTGCTACATCAGTAGTGGTGGTGGATCGTAGCAACGAGG 581
Qy 671 GCATTGCTATTGTCACCTTCAGTTCACATATTTATGGTAAATCTGTAAGAACTG 730
Db 582 GCATTGCCATATTCGCTCGGCTGAGTTCACCTTCCTGTTGGTGGCTGAGTGAAGCTG 641
Qy 731 GGTGAGTTTTGGACAATGTCGTGCTTATCCTATTCTATATGTTCTGCTTGGG 790
Db 642 GATCGGTGTTGTTGGCGCGCAGCGCTTTGCTCTACTCTCATGCTGGTGGCGCTGG 701
Qy 791 GTGGTATTGATTTATCATCAATCTTATTCACCTGCACTGATTTGTTGTTACTGATCG 850
Db 702 GTGGCTAGCTGTTGATCATCACTGATACCCCTGACGCTTCCTGACTGCTCATTTAG 761
Qy 851 AGAGTACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTCATGTTGGGTTTAATAT 910
Db 762 GCAGTACTCGCGCGTCTGCTGACCACTACAGCACCTTCTACATCCTGGGACTGCTGT 821
Qy 911 TATCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTGAACATGCGCAG 970
Db 822 TCTCCATGCGATCCCTTCTGGGATTCACCGATACGCAACAGTGAACATGCGCTG 881
Qy 971 CTGAGGTGCTTTGCAATGCTGCAAGTTATGCTTTCTGCAAGTATCTGAGAGCCGAT 1030
Db 882 CGCTGGGAGTGTGCTGCTTATGCGCGTGGCCACCTTGGCCCATTTGCACTCGCTGC 941
Qy 1031 TAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGATCACTAGCTGCAAGTGGCTG 1090
Db 942 TGTGCGCAACAGATTCGGAAGCTGTTTCATCGTGGCGGATTTGCTGGTGGCGTTGGCG 1001
Qy 1091 TCTTCTTAGTGTCTATTTGATTTATACAGTTACATGCACTGAGTGGAGTGGAGT 1150
Db 1002 TCTTTGGCGCTCGTGGTCTACCATGCTGGCGGTGTTGGCCCGCTGGAGTGGAGCT 1061
Qy 1151 TTATTTCATTTGGGATATCGGTATGCAAAATATACATTTCCAAATTTATGATCAGTGT 1210
Db 1062 TCTACTCGCTGGGATATCTGGTACGCAAGATCCACATTTCCCATTTGATCGCTGT 1121
Qy 1211 CTGAGCATCACTTACGATTTGGTGTCTTCTTTTGTATCTACATTTCTTGTATGTA 1270
Db 1122 CGAGCATCAGCCCACTTGGTGTCTGTTCTTCTTGTATCTGCACTCTGCTGGTGGG 1181
Qy 1271 CCTTCCCGAGCGCTTTGGTCTGCAATCAAAATATCAACGATGAAGAGTATTTGTTG 1330
Db 1182 CCTTCCCGAGTGGTGTGCTGCAATCAAGCAGATCAACGAGCGCGCTTTTCTGTTG 1241
Qy 1331 CTCTATATGCAATCAGTCTCTTCTTCTTCTGAGTGTGAGTGGCTGATGTTGACTT 1390
Db 1242 TGTCTGTAGCCATCAGTGGCTTACTTCTGCTGTTGATGTTGGTGGTGTGATGTTGACCC 1301
Qy 1391 TGACTCCAGTCTGTGATGCTGTGCAATTTGCTTTTCAAAATTTTGTAGCACTATT 1450
Db 1302 TCACGCGGTGTGCTGATGCTGGCGGAGTGGCTTTTCCGGACTGTGGATGTGCTCC 1361
Qy 1451 TGGGGGATGA-----CATGAAGAGGAATCCACTGTGGAG-----GACAGCAGT 1498
Db 1362 TGCAAGAGATTCTGTGAAGGATGGCAGCAGCCTAAGCGAGCGCCAGTAGTGGATG 1421
Qy 1499 ATGAGGATGACAAAAGAACCAAGAAATTTGTTATGAAGGAGGTTAAAGTGAAGAAAC 1558
Db 1422 AAGCTGAGGATTTCCATTGAGAAGAGAGCTGTGACGACAGGCTGGCAAGCTGAAGCATC 1481
Qy 1559 ATGCAACTGAACAGGAAAGAACTGAAGAGGATTAAGCCCTTAATATAAAGGATTTGCA 1618
Db 1482 GTACTAAGCATGATCCCGACAGAGTACTGCGTCACTCACTCAAGAGTATTTGTTA 1541
Qy 1619 CCAATGATGCTGATGATGATGATGTTTGTGCTCACTGATGCTGGTCAAGCA 1678

Db 1542 TTTTGGCGGTCTATGCTGTTGATGATGTTGCTGCTGCCACTGCACGTGGGTGACCAAGCA 1601
Qy 1579 ATCCCTACTACTAGTCCAAAGTGTAGTCTGGCTCATACAA---TCATGATGGACCAAGCA 1735
Db 1602 ATGCCTACTCCAGTCCCTCCATTTGCTTGGCTTTTCCACACAGTCAAGATGGATCCCGCA 1661
Qy 1736 ATATCTTAGATGATTTTAGAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATG 1795
Db 1662 ACATTTAGACGATTTAGAGAGGCTTACTACTGGCTTTCCGACACACTGCGGATGATG 1721
Qy 1796 CACGAGTAAATGCTTGGTGGGATTTAGCTATCAGATAGCTGGAATGCTTAATAGAACTA 1855
Db 1722 CTCGCTGATGCTTGGTGGGATTTAGCGATACAGATAGCGGGAATGCAAAACAGACGA 1781
Qy 1856 CGTTGTTGATTAATACACCTCGAATTAACAGCCACATAGCACTGCTGGGAAAGCTATGT 1915
Db 1782 CGCTAGTGGATTAATAGCTGGAACATAGTCAATAGCCTGTTGGCAAGCAATGT 1841
Qy 1916 CTTCTAATGAACAGCAGCCTTAATAATCATGAGGACTCTAGATGTAGATTTATTTTGG 1975
Db 1842 CTTCAACCGAGGAGAGTCTTACGAAATTTATGACATCTCTTGGCTGGACTACGTTTGG 1901
Qy 1976 TTATTTTGGAGGGTATTGCTTATTTGCTGATGATATCAACAAATTTCTCTGATGG 2035
Db 1902 TGAICTTTGGCGGTGATCGGCTATTTCTGGGATGATATCAACAGTCTCTGTTGGATG 1961
Qy 2036 TTAGGATAGTGAAGGAGAACTCCCAAGACATTCGGGAAAGTGAATTTTACCCAC 2095
Db 1962 TCGAAATTTCTGAGGAGAGCATCCCAAGGACATTTAAGAAAGGATTTACTTTACCGACC 2021
Qy 2096 AGGAGAAATTCGTTAGACAAAGCAGGATCCCTTACTTTTGAATTTGCTTATGTATA 2155
Db 2022 GCGTGAATTCAGGTTAGATGCCGAGTGTCTCGGCCCTGCTCAACTGCTTATGTACA 2081
Qy 2156 AATGTCATACACTACAGATTTGAGAAATGCACTGCTGATTTTCTGACACCCAGCTTTG 2215
Db 2082 AATTAAGTACTACAGATTTGAGGAAATGAAGTTGGACTACAGAGTCCATCTGGATG 2141
Qy 2216 ACCGAACACGATATGCTGAGATTTGAAATTAAGGACATTAATAATCAACATTTGGAAGA 2275
Db 2142 ATCGCACAGTAACGCGCTCATTTGGAAATAGGACTTCGATCTGACCTACCTCGAGGAGG 2201
Qy 2276 CTTTACATCAGACACTGCTGTTAGGATATATAAAGTAAAGACCTGATACAGGG 2335
Db 2202 CTTACACACAGAACTGCTGTTTCCATCTAGGTTGAAGAGCCGCTGATGATCA 2261
Qy 2336 AGACATTAGATCACAAACCTCGAGTCACCAACATTTCCCAAAACAGAGTATTTGTCAA 2395
Db 2262 ATAGACCATCACTGAAGACCAAGAGAGAGAGATTCCTCCAGCAACTTCATTTCCAGAA 2321
Qy 2396 AGAGACTACCAAGAGAGCTGCTACATTAATAAATAGCTGTTTAAAGAAAGCA 2455
Db 2322 AGAATCTTAAGCTGCAAG---GGCTACATACGAAACCGACCGGTTGTTGTTAAGGAA 2378
Qy 2456 AGAAATATCTAAGAGAC 2474
Db 2379 AAGCAACCTTGAATAAAC 2397

RESULT 13

CQ589353

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

CQ589353 Sequence 17111 from Patent WO0171042.
CQ589353 GI:41648215

2699 bp DNA

linear

PAT 02-FEB-2004

Drosophila sp.

Drosophila sp.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1

Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.

1062 TCTACTCGCTGGGATGACTGGCTACGCCAAGATCCACATTCCTCCATTCATTCGATCCGTTG 1121
 1211 CTGACATCAACCTACGACTGGGTGCTTCTTCTTGATCTACATATTCCTTGATGTA 1270
 1122 CGGACATCAGCCACACACTGGTCTCTGCTTCTTCTTGATCTGCACTCTGCTGGTGG 1181
 1271 CTTCCAGAGAGGCTTGGTCTGCAATCAAAATATCAACGATGAAGAGTATTTGTTG 1330
 1182 CTTCCAGAGGAGTGGTGGTCTGCTACATCAAGCAGATCAACGAGAGCGCTTTTCGTTG 1241
 1331 CTCTATATGCAATCAGTCTCTCTACTTCTGAGTGGTGGTGGTGGTGGTGGTGGTGGT 1390
 1242 TCTCTAGCCATCAGTGGCTTACTTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1301
 1391 TGACTTCAGTCTGCTGATGCTGCTGCAATTCGCTTTTCAATGTTTTTGAGCACTATT 1450
 1302 TCACGCCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1361
 1451 TGGGGGATGA-----CATGAAAGGGAATCCACCTCTGGAG-----GACAGCAGTG 1498
 1362 TGCAAGAGGATCTCTAAGCGATGGGACAGCCATAGCGCAGCCAGCGAAGTGGATG 1421
 1499 ATGAGGATGACAAAGAAACCAAGGAAATTTGATGATGAGGAGGTAAGTGGAGAAAC 1558
 1422 AAGCTGAGGATTCATTTGAGAGAGAGCGCTGTACGACAGGCTGGCAAGCTGAAGCATC 1481
 1559 ATGCAACTGAACAGGAAACTCAAGAGGATAGGCCCTATATAAAAGCATTTGCA 1618
 1482 GTACTAAGCATGATGCCAGCAGATCTGGCTGAGTCCACCTGAAGATTTGTTA 1541
 1619 CCAATTTGATGCTGATGCTATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1678
 1542 TTTTGGCGGTTCTAATGCTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1601
 1679 ATGCTACTCTAGTCAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1735
 1602 ATGCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1661
 1736 ATATCTTAGATGATTTAGAGAGCTTACTTTTGGCTAGGCAAAATACAGATGACATG 1795
 1662 ACATTTAGAGATTTAGAGAGCTTACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1721
 1796 CACGAGTAACTCTGCTGGGATTTATGGCTATCAGATAGTGGATGGCTAATAGAACTA 1855
 1722 CTCGGTTATGCTTGTGGGATTAACGATACAGATAGCGGATGGCAATGGCAACAGACGA 1781
 1856 CGTTGGTGAATAACACCTGGAATTAACAGCAGATAGCAGTGGTGGGAAAGCTATGT 1915
 1782 CGCTAGTGGATTAATACGTGGAACTAGTACATAGCGCTGGTGGCAAGGCAATGT 1841
 1916 CTTCTAATGAACAGCAGCTTAAATCATGAGGACTCTAGATGTAGATTTATGTTTTGG 1975
 1842 CTTCAACCGAGAGAGTCTTACGAAATTAATGATCTCTGACGTGGACTACGTTTTGG 1901
 1976 TTATTTTGGAGGGTATTTGGCTATTTCTGCTGATGATACAAATTTCTCTGATGG 2035
 1902 TGATCTTTGGCGGTGATCGGCTATTTCTGCGATGATACAAATTTCTGATGG 1961
 2036 TTAGATAGTGAAGAGAAACATCCCAAGAGATTCGGGAAAGTACTATTTTACCCAC 2095
 1962 TCCGAATTTCTGAGGAGAGATCCCAAGAGATTAAGGAAAGCGATTAATTTTACCGACC 2021
 2096 AGGAGAGATTCGGTGTAGCAAAACAGGATCCCTACTTTTGTGATTCCTTTAGTATA 2155
 2022 GCGGTGAATCAGGTGATGCGAAGTGTCTCGGCGCTGCTCACTGCTCTTATGTACA 2081
 2156 AAATGCTATCTACAGATTTGGAGAAATGAGTGGATTTTCTGACACCCCGAGTTTGG 2215
 2082 AATTAAGTACTACAGATTTGGGAAATTAAGTGGATTTTCTGACAGAGTCCATCTGATATG 2141
 2216 ACCGAACAGTAACTGCTGAGATTTGGAATAAGGACATTAATTAACAAATTTGGAGAG 2275
 2142 ATCGCACACGTAAACCGCGCTCATTGGGAATTAAGGACTTCGATCTGACTCTGCTGGAGG 2201

2276 CTTTACATCAAGAACACTGGCTTGTAGGATATATAAAGTAAAGACACCTGTATAACAGG 2335
 2202 CCTACACCACAGAACACTGGCTTGTTCGATCTATAGGTTGAAGAGCCGATGATTTCA 2261
 2336 AGACATAGATCAAAACTCGAGTACCAACATTTTCCCAAAACAGAGATTTTGTCAA 2395
 2262 ATAGACCATCACTGAAGACCAAGGAGAGAGATTTCTCCAGCAAACTTCATTTCAGAA 2321
 2396 AGAAGACTACCAAAAGAGGCTGGCTACATTAATAAATAAGCTGGTTTTTAAGAAAGGCA 2455
 2322 AGACTCTAAGCGTGGCAAG--GGCTACATACGAAACCGGCTGTTGTTAAGGAA 2378
 2456 AGAATATCTTAAGAGAC 2474
 2379 AACGACCTTGAATAAAT 2397

RESULT 12
 AX799088
 LOCUS 2417 bp DNA linear PAT 08-OCT-2003
 DEFINITION Sequence 7 from Patent WO03054008.
 AX799088
 ACCESSION AX799088.1 GI:37605061
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
 1 Ferrel, C. and McBride, K.
 Mammalian SIMP protein, Gene sequence and uses thereof in cancer
 therapy
 Patent: WO 03054008-A 7 03-JUL-2003;
 Compatigene Inc. (CA)
 Location/Qualifiers
 1..2417
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /note="STT3, Acc# AF132552"

JOURNAL
 TITLE
 FEATURES
 source
 1..2417
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /note="STT3, Acc# AF132552"

ORIGIN
 Query Match 38.5%; Score 955.8; DB 6; Length 2417;
 Best Local Similarity 64.8%; Pred. No. 8.7e-179;
 Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;
 191 AGCGGCTGGGTGGGAGTGGCTTCTCTTCACATCCTCTTCCTGGCTGGCTGGCTGGCG 250
 102 AGGTGGCTGGCTACAGAGCCCTAATCACTTCGCCATCCCTTCGCTAATCGCTGGCTGGCG 161
 251 GCTTCAGCTCGGCTCTTCGCCCTTCATCCGCTTCGAAAGCATCATCCAGAGTTCGAC 310
 162 GATTTTCTCTCGCTCTTCGCCGCTATCCGTTTCGAGTCGATTCATGAGTTGATC 221
 311 CGTGGTTTAACTATAGATCAACATCATCTTCGATCTCATGGTCTCTAGATTTTAA 370
 222 CGTGGTTTAACTACCGGCCACCGCTACATGTTGGTCTGCTGCTGCTGCTGCTGCTGCT 281
 371 ATTGGTTTATGAAGAGAGATGGTATCCATCAGGAGAAATAGTAGTGGTGGTGGTGGT 430
 282 ACTGGTTTCAGCGCGCATGGTATCGCTCGCAGGATTTGGGCGGTACCGTCTATC 341
 431 CAGGGTTGATGATAACCGCTGGCTTATTCATTCGATTTTAAATACATTAACATTAAC 490
 342 CCGGCTGATGATTAACGTTCCGCGGAAATCCATTTGGTCTGCTGCTGCTGCTGCTGCTGCT 401
 491 TTCACATAGAGAGATGATGTTGTTCTTCGACCAACTTTTAGCGGCTTTACATCTATAT 550
 402 TCCATATTCGTGACATCTGCGTGTCTGCGCGGATCTTCAGTGGGCTGACCTCCATCT 461
 551 CTACTTTCTGCTTACAAGAGAACTTTTGAACCAAGGAGGAGGACTTTTAGCTGCTTGT 610

QY 2050 GGAGAACATCCCAAGACATTCGGAAAGTACACTATTTTACCCACAGGAGAAATTCGGT 2109
 DB |||||
 QY 541 GGAGAACATCCCAAGACATTCGGAAAGTACACTATTTTACCCACAGGAGAAATTCGGT 600
 DB |||||
 QY 2110 GTAGACAAAGCAGGATCCCTTACTTTGTTGAATTCGCTTATGTATATAAATGTCTACTAC 2169
 DB |||||
 QY 601 GTAGACAAAGCAGGATCCCTTACTTTGTTGAATTCGCTTATGTATATAAATGTCTACTAC 660
 DB |||||
 QY 2170 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCAGCGACAGGTAAT 2229
 DB |||||
 QY 661 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCAGCGACAGGTAAT 720
 DB |||||
 QY 2230 GCTGAGATTGGAATTAAGGACATTAATTCACACATTTGGAAGACCTTTTACATCAGAA 2289
 DB |||||
 QY 721 GCTGAGATTGGAATTAAGGACATTAATTCACACATTTGGAAGACCTTTTACATCAGAA 780
 DB |||||
 QY 2290 CACTGGCTGTTCGATATATAAGTAAAGACACCTGATACACAGGAGACATTAATCAG 2349
 DB |||||
 QY 781 CACTGGCTGTTCGATATATAAGTAAAGACACCTGATACACAGGAGACATTAATCAG 840
 DB |||||
 QY 2350 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAA 2409
 DB |||||
 QY 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAA 900
 DB |||||
 QY 2410 AGGAGCTGTCTACATTAATAAATTAAGCTGGTTTAAAGAGGCAAGAAATATCTAG 2469
 DB |||||
 QY 901 AGGAGCTGTCTACATTAATAAATTAAGCTGGTTTAAAGAGGCAAGAAATATCTAG 960
 DB |||||
 QY 2470 AAGACTGTTTAA 2481
 DB |||||
 QY 961 AAGACTGTTTAA 972
 DB |||||

RESULT 10

BD063986 2546 bp DNA linear PAT 27-AUG-2002
 LOCUS Secreted proteins and polynucleotides encoding them.
 DEFINITION
 ACCESSION BD063986
 VERSION BD063986.1 GI:22609589
 KEYWORDS JP 2001506848-A/10.
 SOURCE unidentified
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 2546)
 AUTHORS Jacobs,K., McCoy,J.M., Lavallie,E.R., Racie,E.A., Verberg,D., Treacy,M., Spaulding,V. and Agostino,M.J.
 TITLE Secreted proteins and polynucleotides encoding them
 JOURNAL Patent: JP 2001506848-A 10 29-MAY-2001;
 GENETICS INSTITUTE INC
 COMMENT PN JP 2001506848-A/10
 PD 29-MAY-2001
 PF 12-DEC-1997 JP 1998525996
 PR 13-DEC-1996 US 08/766263
 PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI DAVID VERBERG,
 PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
 C12N15/12, C12N5/10, C07K14/47, A61K38/17
 CC Strandedness: Double;
 CC Topology: Linear;
 FH Key Location/Qualifiers.

FEATURES
 source
 1..2546
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"

ORIGIN

Query Match 39.2%; Score 972; DB 6; Length 2546;
 Best Local Similarity 100.0%; Pred. No. 5.4e-182;
 Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1510 AAAAGAAACCAAGGAAATTTGTATCATAAAGCAGTAAAGTGAAGGAAACATGCAACTGAA 1569

DB 1 AAAGAAACCAAGGAAATTTGTATCATAAAGCAGTAAAGTGAAGAAACATGCAACTGAA 60
 QY 1570 CAGGAAAACTGAGAGGGATAGCCCTTAATATAAAAGCANTGTCCACATGTTGATG 1629
 DB 61 CAGGAAAACTGAGAGGGATAGCCCTTAATATAAAAGCANTGTCCACATGTTGATG 120
 QY 1630 CTGATCTATTGATGATGTTTCTCTCCACTGTACTCTGGGTCAACAAGCAATGCTACTCT 1689
 DB 121 CTGATCTATTGATGATGTTTCTCTCCACTGTACTCTGGGTCAACAAGCAATGCTACTCT 180
 QY 1690 AGTCCAAAGTGTAGTCTGCTCATACAATCATGATGSCACAGGAATATCTTAGATGAT 1749
 DB 181 AGTCCAAAGTGTAGTCTGCTCATACAATCATGATGSCACAGGAATATCTTAGATGAT 240
 QY 1750 TTTAGAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTCT 1809
 DB 241 TTTAGAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTCT 300
 QY 1810 TGTGGATATGCTATCAGATAGCTGGAATGGCTAATAGAACTAGTTGGTGGATAAT 1869
 DB 301 TGTGGATATGCTATCAGATAGCTGGAATGGCTAATAGAACTAGTTGGTGGATAAT 360
 QY 1870 AACACCTGGAATAACAGCCACATAGCACTGGTGGGAAAGCTATGCTCTTAATGAACA 1929
 DB 361 AACACCTGGAATAACAGCCACATAGCACTGGTGGGAAAGCTATGCTCTTAATGAACA 420
 QY 1930 GCAGCTTAAATCATGAGGACTCTAGATGATGATGTTTGGTATTTTGGAGGG 1989
 DB 421 GCAGCTTAAATCATGAGGACTCTAGATGATGATGTTTGGTATTTTGGAGGG 480
 QY 1990 GTTATTGGCTATTCTGGTATGATATCAACAAATTTCTCTGATGGTTAGGATAGCTGAA 2049
 DB 481 GTTATTGGCTATTCTGGTATGATATCAACAAATTTCTCTGATGGTTAGGATAGCTGAA 540
 QY 2050 GGAGAACATCCCAAGACATTCGGAAAGTACACTATTTTACCCACAGGAGAAATTCGGT 2109
 DB 541 GGAGAACATCCCAAGACATTCGGAAAGTACACTATTTTACCCACAGGAGAAATTCGGT 600
 QY 2110 GTAGACAAAGCAGGATCCCTTACTTTGTTGAATTCGCTTATGTATATAAATGTCTACTAC 2169
 DB 601 GTAGACAAAGCAGGATCCCTTACTTTGTTGAATTCGCTTATGTATATAAATGTCTACTAC 660
 QY 2170 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCAGCGACAGGTAAT 2229
 DB 661 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCAGCGACAGGTAAT 720
 QY 2230 GCTGAGATTGGAATTAAGGACATTAATTCACACATTTGGAAGACCTTTTACATCAGAA 2289
 DB 721 GCTGAGATTGGAATTAAGGACATTAATTCACACATTTGGAAGACCTTTTACATCAGAA 780
 QY 2290 CACTGGCTGTTCGATATATAAGTAAAGACACCTGATACACAGGAGACATTAATCAG 2349
 DB 781 CACTGGCTGTTCGATATATAAGTAAAGACACCTGATACACAGGAGACATTAATCAG 840
 QY 2350 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAA 2409
 DB 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAA 900
 QY 2410 AGGAGCTGTCTACATTAATAAATTAAGCTGGTTTAAAGAGGCAAGAAATATCTAG 2469
 DB 901 AGGAGCTGTCTACATTAATAAATTAAGCTGGTTTAAAGAGGCAAGAAATATCTAG 960
 QY 2470 AAGACTGTTTAA 2481
 DB 961 AAGACTGTTTAA 972
 DB |||||

RESULT 11

AF132552 2417 bp mRNA linear INV 12-FEB-2003
 LOCUS Drosophila melanogaster GM01839 full insert cDNA.
 DEFINITION
 ACCESSION AF132552

Db 301 GGTGCTTTCTTCTTCTGATCTACATATTTCTGTATGTACCTTCCAGAGGCTATGTT 360
Qy 1293 CTGCATCAAAAATATCAACGATGAAGAGATTTTGTCTATATGCAATCAGTCTGT 1352
Db 361 CTGCATCAAAAATATCAACGATGAAGAGATTTTGTCTATATGCAATCAGTCTGT 420
Qy 1353 CTACTTTGCTGAGTGTGTGCGAGTGTGTGACTTTGACTTCCAGTGTGTGTATGCT 1412
Db 421 CTACTTTGCGGATGTGTGTGCGGCTGTGTGACTTCCAGTGTGTGTATGCT 480
Qy 1413 GTCTGCAATTCCTTTTCAATGTTTTTGTGACTATTTTGGGGATGACATGAAAAGGGA 1472
Db 481 GTGCGCATGCTTCTCCTATGTTTTTGTGACTATTTTGGGGATGACATGAAAAGGGA 540
Qy 1473 AAATCCACCTGTGAGGACAGCTGTGATGAGGATGACAAAAGAAACCAAGAAATTTGTA 1532
Db 541 AAATCCACCTGTGAGGACAGCTGTGATGAGGATGACAAAAGAAACCAAGAAATTTGTA 600
Qy 1533 TGATAAGCAGGTAAAGTGAGGAACATGCAACTGACAGGAAACCACTGAAGAGGGATT 1592
Db 601 TGACAAGCAGGTAAAGTGAGGAACATGCAACTGACAGGAAACCACTGAAGAGGGATT 660
Qy 1593 AGGCCCTAATATAAAGCAATGTCACCATGTTGTGCTGTGATGCTATGATGATGCTTTC 1652
Db 661 GGGCCCCAATCAAAAGCATGTGACCATGCTGTGCTGATGCTGATGCTGCTGCTGCTGCT 720
Qy 1653 TGTCCACTGTACCTGGGTCAAGCAATGCTACTCTAGTCCAAAGTGTAGTCTGCTGCTC 1712
Db 721 GGTCCACTGTACCTGGGTCAAGCAATGCTACTCTAGTCCAAAGTGTAGTCTGCTGCTC 780
Qy 1713 ATCAATCATGATGCGACCGAGGATATCTTAGATGATTTTGAAGAGCTTACTTTGGCT 1772
Db 781 CTACATCATGATGCGACCGAGGATATCTTAGATGATTTTGAAGAGCTTACTTTGGCT 840
Qy 1773 AAGCAAAATACAGATGAACATGCAAGTAAATGCTTGTGTGGATTTATGCTATCAGAT 1832
Db 841 GAGCAAAATACAGATGAACATGCAAGTAAATGCTTGTGTGGATTTATGCTATCAGAT 900
Qy 1833 AGCTGGATGCTAATAGACTAGTGTGTGGATATACACCTGGAATTAACAGCCACAT 1892
Db 901 TGCTGGATGCTAATAGACTAGTGTGTGGATATACACCTGGAATTAACAGCCACAT 960
Qy 1893 AGCACTGTGGAAAGCTATGCTTCTAATGAACAGAGGCTATATAAATCATGAGAC 1952
Db 961 CGCACTGTGGAAAGCTATGCTTCTAATGAACAGAGGCTATATAAATCATGAGTC 1020
Qy 1953 TCTAGATGATATGTTTTGTTTATTTTGTGAGGGTTATGCTGATTTCTGTGATGA 2012
Db 1021 CTTTGAATGATATGTTTTGTTTATTTTGTGAGGGTTATGCTGATTTCTGTGATGA 1080
Qy 2013 TATCAACAAATTTCTGTGATGTTTGGATAGTCTGAAGGAGAACATPCCAAAGACATTCG 2072
Db 1081 TATCAACAAATTTCTGTGATGTTTGGATAGTCTGAAGGAGAACATPCCAAAGACATTCG 1140
Qy 2073 GGAAGTGAATTTTATCCCAAGGAGAAATTCCTGTAGACAAAGAGGAGTCCCTTAC 2132
Db 1141 GGAAGTGAATTTTATCCCAAGGAGAAATTCCTGTAGACAAAGAGGAGTCCCTTAC 1200
Qy 2133 TTTGTTGAATTCCTTATGATATAAATGTCATACATGATTTGGAGAAATGCAGCTGGA 2192
Db 1201 TCTGTTAACTGCTTATGATATAAATGTCATACATGATTTGGAGAAATGCAGCTGGA 1260
Qy 2193 TTTTGTGATACCTTGTGAGGAGAAATGCTGATGATTTGGAGAAATGCAGCTGGA 2252
Db 1261 TTTTGTGATACCTTGTGAGGAGAAATGCTGATGATTTGGAGAAATGCAGCTGGA 1320
Qy 2253 TAAATTCAAATTTGGAAGAGCTTTTATCATGAAACACTGCTGTTGTTAGGATATATA 2312
Db 1321 TAAATTCAAATTTGGAAGAGCTTTTATCATGAAACACTGCTGTTGTTAGGATATATA 1380
Qy 2313 AGTAAAGCAGCTGATTAACAGGAGACATTTAGATCAAAACCTCGAGTCAACCAATTTT 2372
Db 1381 AGTAAAGCAGCTGATTAACAGGAGACATTTAGATCAAAACCTCGAGTCAACCAATTTT 1440

Qy 2373 CCCAAACAGAAAGTATTTGTCAAGAAGACTACCAAAAGGAGGCTGCTACATTAATAA 2432
Db 1441 CCCAAACAGAAAGTATTTGTCAAGAAGACTACCAAAAGGAGGCTGCTACATTAATAA 1500
Qy 2433 TAAGCTGCTTTTAAAGAAAGCAAGAAATATCTTAAGAGACTGTTTAA 2481
Db 1501 TAAGCTAGTGTTTAAAGAAAGCAAGAAACCTCTTAAGAAAGACTGTTTAA 1549
RESULT 9
AX099510 2546 bp DNA linear PAT 02-APR-2001
LOCUS
DEFINITION
Sequence 150 from Patent WO0119988.
AX099510
VERSION
AX099510.1 GI:13538588
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
Jacobs, K.; McCoy, J.M.; Lavallie, E.R.; Collins-Racie, L.A.; Evans, C.;
Merberg, D.; Treacy, M.; Bowman, M.R.; Spaulding, V. and Agostino, M.J.
TITLE
Secreted proteins and polynucleotides encoding them
JOURNAL
Patent: WO 0119988-A 150 22-MAR-2001;
Genetics Institute, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2546
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
Query Match 39.2%; Score 972; DB 6; Length 2546;
Best Local Similarity 100.0%; Pred. No. 5,4e-182;
Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1510 ABAAGAAACCAAGGAAATTTGTATGATAGGCAAGTAAAGTGAGGAAACATGCAACTGAA 1569
Db 1 ABAAGAAACCAAGGAAATTTGTATGATAGGCAAGTAAAGTGAGGAAACATGCAACTGAA 60
Qy 1570 CAGGAAAAAATCAAGAGGATTTAGGCCCTAATATAAAGCAATGTCACCATGTTGATG 1629
Db 61 CAGGAAAAAATCAAGAGGATTTAGGCCCTAATATAAAGCAATGTCACCATGTTGATG 120
Qy 1630 CTGATCTATTGATGATGTTTCTGCTCCACTGTACCTGGSTCACAAGCAATGCGCTACTCT 1689
Db 121 CTGATCTATTGATGATGTTTCTGCTCCACTGTACCTGGSTCACAAGCAATGCGCTACTCT 180
Qy 1690 AGTCCAAAGTGTAGTCTGCTCCTCATACAAATCATGATGCGACCCAGGAATATCTTAGATGAT 1749
Db 181 AGTCCAAAGTGTAGTCTGCTCCTCATACAAATCATGATGCGACCCAGGAATATCTTAGATGAT 240
Qy 1750 TTTTAGAGAGCTTACTTTTGGCTTAAGCAAAATACAGATGAACATGCAAGGATATGCT 1809
Db 241 TTTTAGAGAGCTTACTTTTGGCTTAAGCAAAATACAGATGAACATGCAAGGATATGCT 300
Qy 1810 TGGTGGGTTATGGCTATCAGATAGCTGGAATGCGCTAATAGAACTAGCTTGGTGGATAAT 1869
Db 301 TGGTGGGTTATGGCTATCAGATAGCTGGAATGCGCTAATAGAACTAGCTTGGTGGATAAT 360
Qy 1870 AACACCTGGAATAACAGCCACATAGCTGCTGGTGGGAAAGCTATGCTTCTTAATGAAACA 1929
Db 361 AACACCTGGAATAACAGCCACATAGCTGCTGGTGGGAAAGCTATGCTTCTTAATGAAACA 420
Qy 1930 GCAGCCTATAAATCATGAGGACTCTAGATGATGATTAATGTTTGGTTATTTTGGAGGG 1989
Db 421 GCAGCCTATAAATCATGAGGACTCTAGATGATGATTAATGTTTGGTTATTTTGGAGGG 480
Qy 1990 GTTATTGGCTATCTGGTGGATGATTAACAAATTTCTCTGATGCTTGGATAGCTGAA 2049
Db 481 GTTATTGGCTATCTGGTGGATGATTAACAAATTTCTCTGATGCTTGGATAGCTGAA 540


```

Db 1288 CTTGTATGACCTCCAGCAGGCCCTTGGTCTGATCAAAAATATCAACGATGAAGA 1347
Qy 1321 GTATTCTTGTCT 1332
Db 1348 GTATTCTTGTCT 1359

RESULT 8
LOCUS BC003206 1734 bp mRNA linear ROD 12-NOV-2003
DEFINITION Mus musculus RIKEN cDNA 1300006C19 gene, mRNA (cDNA clone
IMAGE:3587254), partial cds.
ACCESSION BC003206
VERSION BC003206.1 GI:13096823
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1734)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, R., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Udwin, T.B., Toshiyuki, S.,
Schaeetz, T.E., Brownstein, M.J., Udwin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, J., Smallos, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2388257
12477932
2 (bases 1 to 1734)
Strausberg, R.
Direct Submission
Submitted (20-PEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC) Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: http://image.lnl.gov
Series: TRAX Plate: 10 Row: 0 Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.
Location/Qualifiers

```

```

source 1. 1734
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech 11"
/db_xref="taxon:10090"
/clone="IMAGE:3587254"
/tissue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clone) outgrowth infected with the virus MMTV."
/clone_lib="NCI CGAP_Lu29"
/lab_host="DH10B"
/vec="vector: PCMV-SPORT6"
<1. 1734
/gene="1300006C19Rik"
/note="synonym: Simp"
/db_xref="LocusID:68292"
/db_xref="MGI:1915542"
<1. 1549
/gene="1300006C19Rik"
/codon_start=2
/product="1300006C19Rik protein"
/protein_id="AAH03206.1"
/db_xref="GI:13096824"
/db_xref="LocusID:68292"
/db_xref="MGI:1915542"
/translation="GFOPIRTSEHWAAGVFAALQAVAFLOYLRLDLTKOEFOTLFFL
GVSLAAGAVFLSVLYLYTGYIAPSGSFYSLMDTGYAKIHPIIASVSEHQPTVVS
FFDHLILVCFPAGLWFCIKNIDDERFVALAISAVYAGVYVRLMLTLTPVCMV
SAIPNSFHYLGDGMKRENPVEDSDSDKRNPNGLYDKAKGVKRVKHTVEQKEE
GLGPNKISIVTMLMLLMMEFAVETWTNAYSSPVLASVYHNGDTRNLVLDFFEA
FWLQRNDEHARVNMWDGYQIAGNARTLVDNNWNSHIALVGRKMSNETAA
YKIMESLDVYLVIFGVGIVSGDDINDKFLVMVRIAEHGPDKIDREGDYFTQGGFR
VDKAGPTLLNCLMYKSYRFGEMQLDFTPGFDRTRNAEIGNDKIKFHLUEEAT
SEHWLAVRKYKAPDNRETLGKPRVINIVPKQKYLKSKTKRKGVYKVKLVFKXGK
KTSKATV"
misc_feature 1. 1249
/gene="1300006C19Rik"
/note="STT3; Region: Oligosaccharyl transferase STT3
subunit. This family consists of the oligosaccharyl
transferase STT3 subunit and related proteins. The STT3
subunit is part of the oligosaccharyl transferase (Otfase)
complex of proteins and is required for its activity.
Otfase transfers a lipid-linked core-oligosaccharide to
selected asparagine residues in the ER"
/db_xref="CDD:pfam02516"

ORIGIN
Query Match 52.7%; Score 1307.4; DB 10; Length 1734;
Best Local Similarity 90.3%; Pred. No. 1.9e-248;
Matches 1398; Conservative 0; Mismatches 151; Indels 0; Gaps 0;
Qy 933 GGGATTCCAGCAATCAGAACAGTGAACACATGGCAGCTGCAGTCTCTTGCATGCT 992
Db 1 GGGATTCCAGCAATCAGAACAGTGAACACATGGCAGCTGCAGTCTCTTGCATGCT 60
Qy 993 GCAAGCTTATGTTCTTTCAGTATCTCAGAGACCCAGTAAACAAAACAAGAGTTCACAG 1052
Db 61 GCAAGCTTACGCTTTTTCAGTATCTCAGAGACCCGTTGACAAAACAAGAGTTCACAG 120
Qy 1053 CTTTTTTTGGGTGTATCTAGCTGCAGGTGCTGTGTCTTCTAGTGTCTCATCTTT 1112
Db 121 CTTTTTTTGGGTGTCTCCTAGCTGCAGGGCGGTGTGTCTTCTAGTGTCTCATCT 180
Qy 1113 GACTTATCAGGTTTACATTCACCATGGAGTGGCAGGTTTATTCATTGTGGGATCTGG 1172
Db 181 GACTATACAGGTTTATATTCACCATGGAGTGGCAGGTTTATTCATTGTGGGATCTGG 240
Qy 1173 GTATGAAAATACACATTCCTCAATTTTTCATCTAGTGTCTGATGATCTAGTCTTG 1232
Db 241 GTATGAAAATACACATTCCTCAATTTTTCATCTAGTGTCTGATGATCTAGTCTTG 300
Qy 1233 GGTGTCTTTCTTCTTGTATCTACATATTTCTGTATGTACCTTCCACAGCAGCTTGGT 1292

```


QY 2281 ACATCAGAACACTGGCTGTTAGATATATAAGTAAAGACCTGATACAGGGAGACA 2340
 Db |||||
 QY 2348 ACATCAGAACACTGGCTGTTAGATATATAAGTAAAGACCTGATACAGGGAGACA 2407
 Db |||||
 QY 2341 TTAGATCAACACCTCGGTCACCAACATTTTCCAAAACAGAGATTTTGTCAAAGAAG 2400
 Db |||||
 QY 2408 CTAGGTCAACACCTCGGTCACCAACATCGTCCCAACAGAGATTTTGTCAAAGAAG 2467
 Db |||||
 QY 2401 ACTACAAAAGGAGCGTGGCTACATTAATAAATAAGCTGTTTAAAGAAAGCAAGAAA 2460
 Db |||||
 QY 2468 ACTACTAAAGAGCGTGGCTAGTTTAAATAAAGCTAGTGTTTAAGAAGGCAAGAAG 2527
 Db |||||
 QY 2461 ATATCTAAGAGACTGTTTAA 2481
 Db |||||
 QY 2528 ACCTCTAAGAGACTGTTTAA 2548
 Db |||||
 RESULT 5
 LOCUS CQ782554 1664 bp DNA linear PAT 17-MAR-2004
 DEFINITION Sequence 2694 from Patent EP1396543.
 ACCESSION CQ782554
 VERSION CQ782554.1 GI:45502338
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Wakamatsu, A., Sugiyama, T., Negai, K., Kojima, S., Otsuki, T. and
 Koga, H.
 TITLE Primers for synthesizing full length cDNA clones and their use
 JOURNAL Patent: EP 1396543-A 2694 10-MAR-2004;
 Research Association for Biotechnology (JP)
 FEATURES
 source 1..1664
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 28..11362
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAP85831.1"
 /db_xref="GI:45502338"
 /translation="MARPSAPESKHKSLSNPSWGLMALGNSRHGHGPGCAQAKHA
 AGKAPKAPAGLSGSGPAGWQSLSTLFLAWLAGFSRUFPAVIRPESIIHEF
 DPWNYRSTHLSHGFEFLNFWDERAWYPLGRIVGGTVPLMITAGLIHIIWLT
 NITHIRDCVFLAPTFSGLSISTELLRELWNQAGLLAACFIAIPEGYISRSVAG
 SFNDEGIAIFALQPTYYLWVKSVKTSVFMTCCLSYFPMVSAWGVYFIINLPLH
 VEVLLMQRYSKRVYIAYTFYIVGLISMOIPFVGFPQIRTSHEHMAAGVPAIIQAY
 AFLQVLRDLTKQBFOTLFFGLVSLAAGVFLSVIYLTVTGVIAPWSGRFVSLWDTGY
 AKIHPIIASVSEHQPTTWVSPFDLHLVCTFPAGLMFCIKNINDERVGA"
 ORIGIN
 Query Match 53.6%; Score 1328.8; DB 6; Length 1664;
 Best Local Similarity 99.8%; Pred. No. 1.le-252;
 Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATGGCGGAGCCCTGGCCCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGTGG 60
 Db |||||
 QY 28 ATGGCGGAGCCCTGGCCCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGTGG 87
 Db |||||
 QY 61 AGTGCCTCATGCGCTTGGGAAACAGCCGGCAGCGCCACCAACCGGCGCGGCGCCAGTGC 120
 Db |||||
 QY 88 AGTGCCTCATGCGCTTGGGAAACAGCCGGCAGCGCCACCAACCGGCGCGGCGCCAGTGC 147
 Db |||||
 QY 121 GGGCAAGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 180
 Db |||||
 QY 148 GGGCAAGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 207
 Db |||||
 QY 181 GGGCTGTGCGAGCGGGCTGGGTGGCGAGTGGTCTCTCTTCAACATCTCTCTCTCTCTCT 240
 Db |||||

Db 208 GGGCTGTGCGAGCGGCTGGGTGGAGTCGCTTCTCTCTTCAACCATCTCTCTCTCTCTCT 267
 QY 241 TGGCTTGGCGGCTTCACTTGGCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
 Db 268 TGGCTTGGCGGCTTCACTTGGCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 327
 QY 301 GAGTTTCAACCCGCTGTTTAACTATAGATCAACACATCATCTTGTCACTCTCATGGTTCAT 360
 Db 328 GAGTTTCAACCCGCTGTTTAACTATAGATCAACACATCATCTTGTCACTCTCATGGTTCAT 387
 QY 361 GAATTTTAAATTTGTTTGGTGAAGAGCATGGTATCCACTAGGAGAAATAGTAGTGGT 420
 Db 388 GAATTTTAAATTTGTTTGGTGAAGAGCATGGTATCCACTAGGAGAAATAGTAGTGGT 447
 QY 421 ACTGTTTACCCAGGTTGATGATPAAACCGCTGGCTTATTCATTGGATTTTAAATACATG 480
 Db 448 ACTGTTTACCCAGGTTGATGATPAAACCGCTGGCTTATTCATTGGATTTTAAATACATG 507
 QY 481 AACATAACTCTTCCACATAGAGAGCATGTTGTGTTCTCTTGGACCAACTTTTAGCGGCTT 540
 Db 508 AACATAACTCTTCCACATAGAGAGCATGTTGTGTTCTCTTGGACCAACTTTTAGCGGCTT 567
 QY 541 ACATCTATATCTACTTCTCTCTCTTCAAGAGAACTTTTGGAAACCAAGGAGCAGGACTTTTA 600
 Db 568 ACATCTATATCTACTTCTCTCTCTTCAAGAGAACTTTTGGAAACCAAGGAGCAGGACTTTTA 627
 QY 601 GCTGCTTGTGTTTATTTGCTATTGTACAGGCTACATATCTCGGTCACTAGTGGATCTCTT 660
 Db 628 GCTGCTTGTGTTTATTTGCTATTGTACAGGCTACATATCTCGGTCACTAGTGGATCTCTT 687
 QY 661 GATAATGAAGGCAATGCTATTTTGGCTTCACTAGTTCACATATCTTATTTGGGTAAAAATCT 720
 Db 688 GATAATGAAGGCAATGCTATTTTGGCTTCACTAGTTCACATATCTTATTTGGGTAAAAATCT 747
 QY 721 GTAAAAACTGGGTCAGTCTTTTGGCAATGCTGCTGCTTATCTCTTCTATATGTC 780
 Db 748 GTAAAAACTGGGTCAGTCTTTTGGCAATGCTGCTGCTTATCTCTTCTATATGTC 807
 QY 781 TCTGCTTGGGTCGTTATGTTATTTATCAATCTTATTTCCACTGCATGTTATTTGTTG 840
 Db 808 TCTGCTTGGGTCGTTATGTTATTTATCAATCTTATTTCCACTGCATGTTATTTGTTG 867
 QY 841 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACCTTTCTACATTTG 900
 Db 868 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACCTTTCTACATTTG 927
 QY 901 GGTTTTAAATTTATCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAACCAAGTGA 960
 Db 928 GGTTTTAAATTTATCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAACCAAGTGA 987
 QY 961 CACATGCGACCTGCGAGTGTCTTTGATTTGCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1020
 Db 988 CACATGCGACCTGCGAGTGTCTTTGATTTGCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1047
 QY 1021 AGAGACCGATTAACAAAACAGAGTTCAGACCTTTCTTTTGGGTGATATCATAGCT 1080
 Db 1048 AGAGACCGATTAACAAAACAGAGTTCAGACCTTTCTTTTGGGTGATATCATAGCT 1107
 QY 1081 GCAGGTGCTGTGTTCTTCTAGTGTCTATTTTGAATTTACAGGTTTACATTTGCACCATG 1140
 Db 1108 GCAGGTGCTGTGTTCTTCTAGTGTCTATTTTGAATTTACAGGTTTACATTTGCACCATG 1167
 QY 1141 AGTGGCAGGTTTATTTATTTGTTGGGATCTGGGTATGCAAAAATACATTTCCAAATATT 1200
 Db 1168 AGTGGCAGGTTTATTTATTTGTTGGGATCTGGGTATGCAAAAATACATTTCCAAATATT 1227
 QY 1201 GCATCAGTGTCTGAGCATCAACCTACGCTTGGGTGCTTTCTTTCTTTGATCTACATATT 1260
 Db 1228 GCATCAGTGTCTGAGCATCAACCTACGCTTGGGTGCTTTCTTTCTTTGATCTACATATT 1287
 QY 1261 CTTGTATGTACCTTCCAGCAGGCTTTGGTCTCTGATCAAAAATATCAACGATGAAAGA 1320
 Db 1288 CTTGTATGTACCTTCCAGCAGGCTTTGGTCTCTGATCAAAAATATCAACGATGAAAGA 1347

Db 137 AGCGGCTCATGGCTCTGGGAAACAGCGCCACAGCGGCACCATGGGCCCGGAAACCCAGAGC 196
Qy 121 CGGCACAAGCGCGGG 180
Db 197 CGGTCCAGGGGG-----GG 247
Qy 181 GGGGTGTGCGAGCGGGGTGGGTGCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
Db 248 GGTGTGCGAGCGCGGG 307
Qy 241 TGGCTTGGCGGCTTCACTGCGGCTTCTGCGGCTCATGCGCTTCGAAAGCATCATCCAC 300
Db 308 TGGCTTGGCGGCTTCACTGCGGCTTCTGCGGCTCATGCGCTTCGAAAGCATCATCCAC 367
Qy 301 GAGTTGCAACCGGTGGTTAACTATAGATCAACACATCATCTTGCATCTCATGCGTTCTAT 360
Db 368 GAGTTGCAACCGGTGGTTAACTATAGATCAACACATCATCTTGCATCTCATGCGTTCTAT 427
Qy 361 GAATTTTAAATGGTTTGAAGAGAGATGATGCTTCCACTAGGAAGATAGTGGTGT 420
Db 428 GAGTTTCTAAATGGTTTGAAGAGAGATGATGCTTCCACTAGGAAGATAGTGGTGGC 487
Qy 421 ACTGTTTACCAGGGTGGTGAATACCGCTGGGCTTATTCATGGAATTTAAATACATG 480
Db 488 ACCGTTTACCAGGGTGGTGAATACCGCTGGGCTTATTCATGGAATTTAAATACATG 547
Qy 481 AACATACCTTCAATAGAGAGATGATGCTTCCACTAGGAAGATAGTGGTGGT 540
Db 548 AACATAACAGTTTCAATAGAGATGATGCTTCCACTAGGAAGATAGTGGTGGT 607
Qy 541 ACATCTATATCTACTTCT 600
Db 608 ACATCTATATCTACTTCT 667
Qy 601 GCTGCTGTTTATGCTATGACAGGCTACATATCGGTCAGTACGTAGTGGTGGT 660
Db 668 GCTGCTGCTTCTATGCTATGACAGGCTACATATCGGTCAGTACGTAGTGGTGGT 727
Qy 661 GATATGAAGCATTTGCTATTTTGGCACTTCACTATCATATCTATTTATGGTAAATCT 720
Db 728 GATATGAAGCATTTGCTATTTTGGCACTTCACTATCATATCTATTTATGGTAAATCT 787
Qy 721 GTAAATCTGGTCACTTTTGGCAATGCTGCTGCTTATCTTCTATATGCT 780
Db 788 GTGAAGACCGGTCTGTGTTCTGGCAATGCTGCTGCTTGTATTTCTACATGCT 847
Qy 781 TCTGCTGGGTGGTATGATTTATCATCAATCTTATTCACATGATGATTTGTTG 840
Db 848 TCTGCTGGGTGGTATGATTTATCATCAATCTTATTCACATGATGATTTGTTG 907
Qy 841 TTACTGATGAGATACAGAAAGATCTACATAGCATATAGCATTTCTACATG 900
Db 908 CTGCTGATGAGAGGTACAGCAAGAGATCTACATAGCATATAGCATTTGATATG 967
Qy 901 GGTTTAATATTAATCAATGAGATACCTTTTGGGATTCAGCCAACTCAGAAAGAGTAA 960
Db 968 GGTTTAATATTAATCAATGAGATACCTTTTGGGATTCAGCCAACTCAGAAAGAGTAA 1027
Qy 961 CACATGGCAGTCTGAGGTGCTTGTGATGCTGCAAGCTTATGCTTCTGAGTATCTG 1020
Db 1028 CACATGGCAGTCTGAGGTGCTTGTGATGCTGCAAGCTTATGCTTCTGAGTATCTG 1087
Qy 1021 AGAGACCGATTAAACAAACAGAGTCTCAGACCTTTTCTTTTGGGTATCACTAGCT 1080
Db 1088 AGAGACCGATTAAACAAACAGAGTCTCAGACCTTTTCTTTTGGGTATCACTAGCT 1147
Qy 1081 CGAGGTGCTGCTTCTGAGTCTATCTATTTGACTATACAGTTACATTTCCACATG 1140
Db 1148 GCAGGCGTGTGCTTCTGAGTCTATCTATCTGACATACAGTTATATGACCATGG 1207
Qy 1141 AGTGCAGGTTTATCTATCTGAGTCTGAGTATGAGTATGAGTATGAGTATGAGTATG 1200
Db 1208 AGTGCAGGTTTATCTATCTGAGTCTGAGTATGAGTATGAGTATGAGTATGAGTATG 1267

Qy 1201 GCATCAGTCTGTGAGCATCAACTAGACTTGGGTGCTCTTCTTCTTGTATCTACATATT 1260
Db 1268 GCATCAGTCTGTGAGCATCAACTAGACTTGGGTGCTCTTCTTCTTGTATCTACATATT 1327
Qy 1261 CTTGTATGATCTTCCAGCAGCGCTTTGGTCTGCTCAAAAATATCAAGATGAAAGA 1320
Db 1328 CTTGTATGATCTTCCAGCAGCGCTTGGTCTGCTCAAAAATATCAAGATGAAAGA 1387
Qy 1321 GTATTTGTTGCTCTATATGCAATCAGTCTGCTCTTCTTCTGAGTATGCTCGACTG 1380
Db 1388 GTATTTGTTGCTCTATATGCAATCAGTCTGCTCTTCTTCTGAGTATGCTCGACTG 1447
Qy 1381 ATGTTGACTTTGACTTCCAGTCTGCTATGCTGCTGCAATGCTTTTCAAAATGTTTT 1440
Db 1448 ATGTTGACTTTGACTTCCAGTCTGCTATGCTGCTGCTGCTGCTTCTTCAAAATGTTTT 1507
Qy 1441 GAGCAGTATTTGGGGATGACATGAAAGGAAATCCACTGTTGGAGGACAGCAGTAT 1500
Db 1508 GAGCAGTATTTGGGGATGACATGAAAGGAAATCCACTGTTGGAGGACAGCAGTAT 1567
Qy 1501 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAAGGAGGATTAAGTGAAGAAAT 1560
Db 1568 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAAGGAGGATTAAGTGAAGAAAT 1627
Qy 1561 GCAACTGAAACAGGAAACCAAGGAAATTTAGGAGGATTAAGTGAAGAAATGTCACC 1620
Db 1628 GTGACAGAGCAAGAAACCAAGGAAATTTAGGAGGATTAAGTGAAGAAATGTCACC 1687
Qy 1621 ATGTTGACTTCTGATCTATGATGATGTTGCTGCTCACTGCTACTGCTGCTCAAGCAAT 1680
Db 1688 ATGTTGACTTCTGATCTATGATGATGTTGCTGCTCACTGCTGCTCAAGCAAT 1747
Qy 1681 GCTTACTCTAGTCCAGTGTAGTCTGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 1748 GCTTACTCTAGTCCAGTGTAGTCTGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCT 1807
Qy 1741 TTAGATGATTTTAGAGAGCTTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGCA 1800
Db 1808 TTAGATGATTTTAGAGAGCTTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGCA 1867
Qy 1801 GTAATCTCTGTTGGATTTATGCTATCAGATAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
Db 1868 GTATGCTGTTGGATTTATGCTATCAGATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1927
Qy 1861 GTGATTAATAACACTGGAAATTAAGCAACATAGCACTGCTGCTGCTGCTGCTGCTGCTGCT 1920
Db 1928 GTGATTAATAACACTGGAAATTAAGCAACATAGCACTGCTGCTGCTGCTGCTGCTGCTGCT 1987
Qy 1921 AATGAAACAGCAGCTTATAAATCATGAGGATCTAGATGATGATGATGATGATGATGATGAT 1980
Db 1988 AATGAAACAGCAGCTTATAAATCATGAGGATCTAGATGATGATGATGATGATGATGATGAT 2047
Qy 1981 TTTGGAGGCTTATTTGGCTTATTTCTGGTATGATATCAACAAATTTCTTCTGATGCTTAGG 2040
Db 2048 TTTGGAGGCTTATTTGGCTTATTTCTGGTATGATATCAACAAATTTCTTCTGATGCTTAGG 2107
Qy 2041 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCCAAGGAA 2100
Db 2108 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCCAAGGAA 2167
Qy 2101 GAATTCGGTGTAGCAAAAGCAGGATCCCTTACTTTGTTGAATTTGCTTATGATTAATAATG 2160
Db 2168 GAGTTCGGTGTAGCAAAAGCAGGATCCCTTACTTTGTTGAATTTGCTTATGATTAATAATG 2227
Qy 2161 TCATCTACAGATTTGGAGAAATGAGTGGATTTTCGTAACCCCAAGGTTTTCACCGA 2220
Db 2228 TCATCTACAGATTTGGAGAAATGAGTGGATTTTCGTAACCCCAAGGTTTTCACCGA 2287
Qy 2221 ACAGCTAATGCTGAGATTTGGAGAAATGAGTGGATTTTCGTAACCCCAAGGTTTTCACCGA 2280
Db 2288 ACAGCTAATGCTGAGATTTGGAGAAATGAGTGGATTTTCGTAACCCCAAGGTTTTCACCGA 2347

Db	2277	AATGAACGGCGCGCTATAAATCATATGAGGTCCCTTGATGTGCATATATGTGTGGTTATT	2333
QY	1981	TTTGAGGGGTATTTCGCTATTCTCGTGTATGATATCAACAAATTTCTCTGATGGTTAGG	2040
Db	2337	TTCCGAGGAGTGATTGGCTATTCCGGGGACGATATCAACAAGTTTCCTCTGATGTCAGG	2396
QY	2041	ATAGCTGAAGGAGAAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCACACAGGA	2100
Db	2397	ATAGCTGNAAGGGGAGCATCCCAAGACATCCGGGAGGTGACTATTTCAACCAGCAGGA	2456
QY	2101	GAATTCGGTGTAGACAAAGCAGGATCCCTACTCTTTGTTGAAATTCGCTTATGTATAAAATG	2160
Db	2457	GAGTTCCGAGTAGACAAAGCTGGGTCTCCTACTCTGTTTAAACTGCCTTATGTATAAAATG	2516
QY	2161	TCATACTACAGATTGTGAGAAATGACAGCTGGAATTTTCGTACACCCCCAGCGTTTTCACCGA	2220
Db	2517	TCATACTACAGATTGTGAGAAATGACAGCTAGATTTTCGCACTCCCCAGCGTTTGACCGA	2576
QY	2221	ACAGCTAATGCTGAGATTGGAATAAGGACATTAATTAACAATTTGGGAAGAGCCCTTT	2280
Db	2577	ACAGCTAATGCTGAGATTGGAATAAAGACATTAATTAACAATTTGGGAGGAAGCTTTT	2636
QY	2281	ACATCAGAACTACTGGCTTGTGTAGGATATATAAAGTAAAGCACCTGTATAACAGGAGACA	2340
Db	2637	ACATCAGAGCACTGGCTTGTGAGATATATAAAGTGAAGCACCTGTACACAGGAGACA	2696
QY	2341	TTAGATCAAACTCGAGTCACCAACATTTTCCAAAAACAGAGATATTGTTCAAAGAG	2400
Db	2697	CTAGGTCACAACTCGAGTCACCAACATCGTCCCAACAGAGATATTGTTCAAAGAG	2756
QY	2401	ACTACCAAAGGAGCGTGCTACTATAAAAAATAAGCTGGTTTTTAAGAAAGCAAGAAA	2460
Db	2757	ACTACTAAAGGAGCGTGCTAGCTATAAATAAAGTAGTGTGTTAAGAAAGCAAGAG	2816
QY	2461	ATATCTAAGAGAGCTGTTAA	2481
Db	2817	ACCTCTAAGAGAGCTGTTAA	2837
RESULT 4			
AX799084			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 2338			
Conservative			
Mismatches			
Indels			
Gaps			
QY	1	ATGGCGGAGCCCTCGGCCCGGAGACGACAGTGTCTCACTGTCTCCCGTGG	60
Db	77	ATGGCGGAGCCCTCGGCCCGGAGACGACAGTGTCTCCCTCACTGTCTCCCGTGG	136
QY	61	AGTGGCTCTATGGCCCTGGGAAACAGCCGGCACGCGCACTACGGCGCGGCCCGCTGTC	120

Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences (26), 16899-16903 (2002)
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257 PUBLISHED
 12477932
 2 (bases 1 to 4236)
 Strausberg, R.
 Direct Submission
 Submitted (15-MAY-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
 Thomas L. Casavant.
 Web site: <http://genome.uiowa.edu>
 Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
 Bonaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,
 Fishler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
 Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J.,
 Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: Plate: Row: Column: 0
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis. Similarity but not identity to protein.

FEATURES

source

1..4236
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="MGC:64679 IMAGE:6837097"
 /tissue="Brain, enriched mouse brain 12.5dpc"
 /clone_lib="NIH BMAP_F10"
 /lab_host="DH10B"
 /note="Vector: pTX-ASC"
 1..4236
 /gene="130006C19Rik"
 /note="synonym: Simp"
 /db_xref="LocusID:68292"
 /db_xref="WGI:1915542"
 366..2837
 /gene="130006C19Rik"
 /codon_start=1
 /product="130006C19Rik protein"
 /protein_id="AAHS2433.1"
 /db_xref="GI:30851502"
 /db_xref="LocusID:68292"
 /db_xref="WGI:1915542"
 /translation="MAEPAPESKXSSLNSSPWSGLMNLGNRHGHGPGTQASRA
 AAFPGPPGLSGLSQPGNQLSFTILFLANLAGFSRLFAVIRFESIHEDFPW
 FNYSRHLASHGFYFLANWDERAYPLGRVGVTVPLMTAGLILHNLITLNP
 VHRDVCVLAFTPLSGLSFTSLTELLRELNNQAGLLAACFAIIVGYISRSVAGSF
 NEGIAIPALQFTYLVKVSFKTGSFVWCCCLSYFVWVAGGVFVILNIPHFVF
 LLLMQRYSRYAYSTFYVILSNQIPFVGFQPIRTSEHMAAGVFAILLQYAFI
 QYLRDLTKQFOTPLFGLVLAAGVFLSVTLVTVGVIAWGRFVSLMDTVGAKI
 HIPIASVSHQPTVWVSFPFDLHILVCTFPAGLMFCIKINDERVFAVLAISVVE
 AGVMVRLMLTLPVCMNLALAEFNVFHLGDDMKRNPVEDSDSDDEKBNPNLY
 DKAGVKRKHTEKEEPEGKPNIKSIIVTMLMLLMLFPAVCTVTSNAYSFSVVL
 ASYNDGTNRILDDPFAEFWLRQNTDEHARVNSWMDYVYIAGVAGVITLVDNNTWN
 NSHIALGVKXASNEFAAYKIMRSLDVLVIFGVIGYSGDDINKFLMWVRAEGE
 HPKDIREGDTYQGFPRVDKAGSPPTLLNCLMYKSYFVGEMLQDFTPPGPDRTN
 AEIGNKDIKFLHEEAPFSEHNLVRIYKVKAPDNRETILGHKPRVTNIVPKOKYLSKKT
 TKRGRGVVKNKLVPFKGKTSKKTV"

gene

CDS

misc_feature 567..2537
 /gene="130006C19Rik"
 /note="STT3. Region: Oligosaccharyl transferase STT3
 subunit. This family consists of the oligosaccharyl
 transferase STT3 subunit and related proteins. The STT3
 subunit is part of the oligosaccharyl transferase (Oase)
 complex of proteins and is required for its activity.
 Oase transfers a lipid-linked core-oligosaccharide to
 selected asparagine residues in the ER"
 /db_xref="CDD:pfam02516"

ORIGIN

Query Match 83.9%; Score 2081.8; DB 10; Length 4236;
 Best Local Similarity 90.3%; Pred. No. 0;
 Matches 2240; Conservative 0; Mismatches 232; Indels 9; Gaps 1;

QY 1 ATGGCGAGCCCTCGGCGCCCGGAGAGCAAGTCCTCAACTCGTCCCGCTGG 60
 DB ATGGCGAGCCCTCGGCGCCCGGAGAGCAAGTCCTCAACTCGTCCCGCTGG 425
 QY 61 AGTGGCTCATGGCCCTTGGGAAACAGCCGGCAGGCCACACAGGGGCCCGGGGCCAGTGC 120
 DB AGTGGCTCATGGCCCTTGGGAAACAGCCGGCAGGCCACACAGGGGCCCGGGGCCAGTGC 485
 QY 426 AGCGGCTCATGGCTTGGGAAACAGCCGGCAGGCCACACAGGGGCCCGGGGCCAGTGC 180
 DB AGCGGCTCATGGCTTGGGAAACAGCCGGCAGGCCACACAGGGGCCCGGGGCCAGTGC 536
 QY 181 GGCTGTGCGACGCGCGCTGGGTCGCTTCTCTTCACATCTTCCTTCGCTGC 240
 DB GGCTGTGCGACGCGCGCTGGGTCGCTTCTCTTCACATCTTCCTTCGCTGC 596
 QY 537 GGCTGTGCGACGCGCGCTGGGTCGCTTCTCTTCACATCTTCCTTCGCTGC 300
 DB GGCTGTGCGACGCGCGCTGGGTCGCTTCTCTTCACATCTTCCTTCGCTGC 656
 QY 241 TGGCTTGGCGCTTCAGTTCGGGCTCTTCGGCGTCATCCGCTTCGAGAGCATCACC 360
 DB TGGCTTGGCGCTTCAGTTCGGGCTCTTCGGCGTCATCCGCTTCGAGAGCATCACC 716
 QY 597 TGGCTTGGCGCTTCAGTTCGGGCTCTTCGGCGTCATCCGCTTCGAGAGCATCACC 480
 DB TGGCTTGGCGCTTCAGTTCGGGCTCTTCGGCGTCATCCGCTTCGAGAGCATCACC 836
 QY 301 GAGTTCGACCGCTGTTAACTATAGATCAACATCATCTTGCATCTCATGGTTCAT 360
 DB GAGTTCGACCGCTGTTAACTATAGATCAACATCATCTTGCATCTCATGGTTCAT 716
 QY 657 GAGTTCGACCGCTGTTAACTATAGATCAACATCATCTTGCATCTCATGGTTCAT 420
 DB GAGTTCGACCGCTGTTAACTATAGATCAACATCATCTTGCATCTCATGGTTCAT 776
 QY 717 GAGTTCGACCGCTGTTAACTATAGATCAACATCATCTTGCATCTCATGGTTCAT 480
 DB GAGTTCGACCGCTGTTAACTATAGATCAACATCATCTTGCATCTCATGGTTCAT 836
 QY 421 ACTGTTACCGAGGTTGATGATAACCGCTGGCTTATCATTCGATTTAAATACATTG 360
 DB ACTGTTACCGAGGTTGATGATAACCGCTGGCTTATCATTCGATTTAAATACATTG 836
 QY 777 ACCGTTTACCGAGGTTGATGATAACCGCTGGCTTATCATTCGATTTAAATACATTG 540
 DB ACCGTTTACCGAGGTTGATGATAACCGCTGGCTTATCATTCGATTTAAATACATTG 896
 QY 481 AACATACCTGTTTACATAGAGAGAGTATGTGTTCCTTGACCACTTTTAGCGGCTT 600
 DB AACATACCTGTTTACATAGAGAGAGTATGTGTTCCTTGACCACTTTTAGCGGCTT 956
 QY 837 AACATACCTGTTTACATAGAGAGAGTATGTGTTCCTTGACCACTTTTAGCGGCTT 660
 DB AACATACCTGTTTACATAGAGAGAGTATGTGTTCCTTGACCACTTTTAGCGGCTT 1016
 QY 541 ACATCTATATCTACTTCTCTTCAAGAGAACCTTTTGGAAACCAAGAGAGAGCTTTTA 720
 DB ACATCTATATCTACTTCTCTTCAAGAGAACCTTTTGGAAACCAAGAGAGAGCTTTTA 1076
 QY 897 ACATCTATATCTACTTCTCTTCAAGAGAACCTTTTGGAAACCAAGAGAGAGCTTTTA 780
 DB ACATCTATATCTACTTCTCTTCAAGAGAACCTTTTGGAAACCAAGAGAGAGCTTTTA 1136
 QY 601 GCTGCTTGTGTTTATGCTTATGCTTACAGGCTACATATCTCGGTGAGTGGATCTCTT 840
 DB GCTGCTTGTGTTTATGCTTATGCTTACAGGCTACATATCTCGGTGAGTGGATCTCTT 1196
 QY 957 GCTGCTTGTGTTTATGCTTATGCTTACAGGCTACATATCTCGGTGAGTGGATCTCTT 900
 DB GCTGCTTGTGTTTATGCTTATGCTTACAGGCTACATATCTCGGTGAGTGGATCTCTT

Db 781 TCTGTTGGGTTGTTATGTTATATCATCACTTTATTCACATGATGTTGTTG 840
Qy 841 TTACTGATGCGAGATACAGCAAAAGAGTCTACATAGCATATAGCATTTCTACATGTTG 900
Db 841 TTACTGATGCGAGATACAGCAAAAGAGTCTACATAGCATATAGCATTTCTACATGTTG 900
Qy 901 GGTTTAAATATATCAATCCAGATACCTTTTGTGGGATCCAGCAATCAGAAACAAGTAA 960
Db 901 GGTTTAAATATATCAATCCAGATACCTTTTGTGGGATCCAGCAATCAGAAACAAGTAA 960
Qy 961 CACATGGAGCTGAGGTCTTTTGTGATGCTGCAAGCTTATGCTTTCTTCAGTATCTG 1020
Db 961 CACATGGAGCTGAGGTCTTTTGTGATGCTGCAAGCTTATGCTTTCTTCAGTATCTG 1020
Qy 1021 AGAGACCGATTAACAAAACAAGAGTTCAGACCCCTTTTCTTTTGGGTGATCAGTACGT 1080
Db 1021 AGAGACCGATTAACAAAACAAGAGTTCAGACCCCTTTTCTTTTGGGTGATCAGTACGT 1080
Qy 1081 GCAGGTGCTGTTCTTTAGTGTCTATTTGACTTATACAGTTACATGTCACCATGG 1140
Db 1081 GCAGGTGCTGTTCTTTAGTGTCTATTTGACTTATACAGTTACATGTCACCATGG 1140
Qy 1141 AGTGCAGGTTTATTCATTGTTGGGATCTGGGTATGCAAAATACACATTCCAATTAT 1200
Db 1141 AGTGCAGGTTTATTCATTGTTGGGATCTGGGTATGCAAAATACACATTCCAATTAT 1200
Qy 1201 GCATCAGTCTGAGCATCACTACGACTTGGGTGCTTTCTTTCTTTGATCTACATAT 1260
Db 1201 GCATCAGTCTGAGCATCACTACGACTTGGGTGCTTTCTTTCTTTGATCTACATAT 1260
Qy 1261 CTTGTATGACCTTCCAGAGGCTTTGTTCTGCAATCAAAAATACAAAGTAAAGA 1320
Db 1261 CTTGTATGACCTTCCAGAGGCTTTGTTCTGCAATCAAAAATACAAAGTAAAGA 1320
Qy 1321 GTATTTGTTGCTATATGCAATCAGTCTGCTGCTTTGCTGAGTGATGGTGCACGTG 1380
Db 1321 GTATTTGTTGCTATATGCAATCAGTCTGCTGCTTTGCTGAGTGATGGTGCACGTG 1380
Qy 1381 ATGTTGACTTTGACTCCAGTCTGTTGATGCTGCTGCAATTCCTTTTCAAAATGTTTT 1440
Db 1381 ATGTTGACTTTGACTCCAGTCTGTTGATGCTGCTGCAATTCCTTTTCAAAATGTTTT 1440
Qy 1441 GAGCCTATTTTGGGGATGACATGAAAGGGAATCCACTGTGGAGGACAGCTGAT 1500
Db 1441 GAGCCTATTTTGGGGATGACATGAAAGGGAATCCACTGTGGAGGACAGCTGAT 1500
Qy 1501 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGAGGTAAGTGAAGAAACAT 1560
Db 1501 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGAGGTAAGTGAAGAAACAT 1560
Qy 1561 GCAACTGAACAGGAAACCAAGGAGGATAGGAGGATTAATAAAGCAATTTGTCACC 1620
Db 1561 GCAACTGAACAGGAAACCAAGGAGGATAGGAGGATTAATAAAGCAATTTGTCACC 1620
Qy 1621 ATGTTGATGCTGATGCTATGATGATGTTGCTGTCACCTGCTGCTGCTGCTGCTGCT 1680
Db 1621 ATGTTGATGCTGATGCTATGATGATGTTGCTGTCACCTGCTGCTGCTGCTGCTGCT 1680
Qy 1681 GCTTACTCTAGTCCCAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 1681 GCTTACTCTAGTCCCAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Qy 1741 TTAGATGATTTTAGAGAGCTTACTTTTGGCTAGGCAAAATACAGATGACATGACGA 1800
Db 1741 TTAGATGATTTTAGAGAGCTTACTTTTGGCTAGGCAAAATACAGATGACATGACGA 1800
Qy 1801 GTAAATGCTTGGTGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGCTTG 1860
Db 1801 GTAAATGCTTGGTGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGCTTG 1860
Qy 1861 GTGGATTAATACACCTGGATTAACAGCCACATAGACCTGTTGGGAAAGCTATGCTTCT 1920

Db 1861 GTGGATTAATACACCTGGATTAACAGCCACATAGCACCTGGTGGGAAAGCTATGCTTCT 1920
Qy 1921 AATGAACAGCAGCCTATTAATATCATAGGACTCTAGATGTAGATATATGTTGGTTATT 1980
Db 1921 AATGAACAGCAGCCTATTAATATCATAGGACTCTAGATGTAGATATATGTTGGTTATT 1980
Qy 1981 TTTGGAGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTGGATGGTTAGG 2040
Db 1981 TTTGGAGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTGGATGGTTAGG 2040
Qy 2041 ATAGCTGAGGAGAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGA 2100
Db 2041 ATAGCTGAGGAGAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGA 2100
Qy 2101 GAATTCGGTGTAGCAAAAGCAGGATCCCTACTTTTGGTGAATTTGCTTATGATAAATG 2160
Db 2101 GAATTCGGTGTAGCAAAAGCAGGATCCCTACTTTTGGTGAATTTGCTTATGATAAATG 2160
Qy 2161 TCATATCTACAGATTTTGGAGAAATGCGCTGGATTTTCGTACACCCCGAGGTTTGAACCGA 2220
Db 2161 TCATATCTACAGATTTTGGAGAAATGCGCTGGATTTTCGTACACCCCGAGGTTTGAACCGA 2220
Qy 2221 ACAGTAAATGCTGAGATTCGAAATTAAGACATTAATTCAAACATTTTGAAGAGCCTTT 2280
Db 2221 ACAGTAAATGCTGAGATTCGAAATTAAGACATTAATTCAAACATTTTGAAGAGCCTTT 2280
Qy 2281 ACATCAGAACACCTGGCTTGTAGGATATATAAAGTAAAGACACCTGATTAACAGGAGACA 2340
Db 2281 ACATCAGAACACCTGGCTTGTAGGATATATAAAGTAAAGACACCTGATTAACAGGAGACA 2340
Qy 2341 TTAGATCAAAACCTGAGTCCACCAATTTTCCAAACAGAGATTTTGTCAAGAGAG 2400
Db 2341 TTAGATCAAAACCTGAGTCCACCAATTTTCCAAACAGAGATTTTGTCAAGAGAG 2400
Qy 2401 ACTACCAAAAGGAGGCTGCTACATTAATAAATAAGCTGGTTTAAAGAAAGCAAGAAA 2460
Db 2401 ACTACCAAAAGGAGGCTGCTACATTAATAAATAAGCTGGTTTAAAGAAAGCAAGAAA 2460
Qy 2461 ATATCTAAGAGACTGTTTAA 2481
Db 2461 ATATCTAAGAGACTGTTTAA 2481

RESULT 3

BC052433 4236 bp mRNA linear ROD 12-NOV-2003
LOCUS BC052433
DEFINITION Mus musculus RIKEN cDNA 130006C19 gene, mRNA (cdna clone MGC:64679
IMAGE:6837097), complete cds.

ACCESSION

BC052433

VERSION

BC052433.1

KEYWORDS

MGC.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 4236)

AUTHORS

Strauberg, R.L., Reingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhut, N.K.,
Hopkins, R.F., Jordan, H., Moore, S.I., Wang, J., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Sheehy, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abrahamson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,

ORIGIN

Query Match 100.0%; Score 2481; DB 6; Length 2481;
Best Local Similarity 100.0%; Pred No. 0;
Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGAGCCCTCGGCCCCGAGAGAGCAAGTGTGCTCTCAACCTGTCGCCGTGG 60
DB 1 ATGGCGAGCCCTCGGCCCCGAGAGAGCAAGTGTGCTCTCAACCTGTCGCCGTGG 60

QY 61 AGTGGCTCATGGCCCTGGGAAACAGCCGCGCCAGCCACACAGGCGCCGCGCCAGTGC 120
DB 61 AGTGGCTCATGGCCCTGGGAAACAGCCGCGCCAGCCACACAGGCGCCGCGCCAGTGC 120

QY 121 GCGCACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 121 GCGCACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180

QY 181 GGGCTGTGCGAGCCGCGGCTGGTGGCAGTGTCTCTCTCTCAACATCTCTCTCTGGCC 240
DB 181 GGGCTGTGCGAGCCGCGGCTGGTGGCAGTGTCTCTCTCTCAACATCTCTCTCTGGCC 240

QY 241 TGGCTTCCGCGCTTCAGCTCGGCGCTTTCGCGCTCATCGCTTCGAAAGCATCATC 300
DB 241 TGGCTTCCGCGCTTCAGCTCGGCGCTTTCGCGCTCATCGCTTCGAAAGCATCATC 300

QY 301 GAGTTCGACCCGCTGTTTAACTATAGATCAACATCATCTTGCATCTCATGGGTTCTAT 360
DB 301 GAGTTCGACCCGCTGTTTAACTATAGATCAACATCATCTTGCATCTCATGGGTTCTAT 360

QY 361 GAATTTTAAATGGTTTGAAGAGCATGTATCCACTAGGAAGATAGTAGGTGT 420
DB 361 GAATTTTAAATGGTTTGAAGAGCATGTATCCACTAGGAAGATAGTAGGTGT 420

QY 421 ACTGTTTACCAGGGTGGTGAATACCGCTGGGCTTATTCAATGGATTTTAAATACATG 480
DB 421 ACTGTTTACCAGGGTGGTGAATACCGCTGGGCTTATTCAATGGATTTTAAATACATG 480

QY 481 AACATACTGTACATAAGAGAGTATGTGTCTGTCGACCACTTTAGCGGCTT 540
DB 481 AACATACTGTACATAAGAGAGTATGTGTCTGTCGACCACTTTAGCGGCTT 540

QY 541 ACATCTATATCTCTTCTCTTACAGAGAACTTTGGAACCAAGAGCAGGACTTTTA 600
DB 541 ACATCTATATCTCTTCTCTTACAGAGAACTTTGGAACCAAGAGCAGGACTTTTA 600

QY 601 GCTGCTTGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 660
DB 601 GCTGCTTGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 660

QY 661 GATAATGAAGCATTTGCTATTTTGGCACTTCAGTTTCACTATCTATTTATGGGTAAATCT 720
DB 661 GATAATGAAGCATTTGCTATTTTGGCACTTCAGTTTCACTATCTATTTATGGGTAAATCT 720

QY 721 GTAAATCTGGGTGAGTATTTTGAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
DB 721 GTAAATCTGGGTGAGTATTTTGAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGT 780

QY 781 TCTGCTTGGGTGAGTATTTTGAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840
DB 781 TCTGCTTGGGTGAGTATTTTGAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840

QY 841 TTACTGTAGAGATCAGCAAGAGTCTACATAGCATATAGCATCTTCTTACATGTG 900
DB 841 TTACTGTAGAGATCAGCAAGAGTCTACATAGCATATAGCATCTTCTTACATGTG 900

QY 901 GGTTTAATTTATCAATGAGATACCTTTTGGGATTTCCAGCACTCAGCAAGTGAA 960
DB 901 GGTTTAATTTATCAATGAGATACCTTTTGGGATTTCCAGCACTCAGCAAGTGAA 960

QY 961 CACATGGAGCTGAGGTGTCTTGGATTCGTCGAGCTTATGCTTCTTGGAGTATCTG 1020
DB 961 CACATGGAGCTGAGGTGTCTTGGATTCGTCGAGCTTATGCTTCTTGGAGTATCTG 1020

QY 1021 AGAGACCGATTAAACAAAACAGAGTTCAGACCCCTTTTCTTTTGGGTGTATCACTAGCT 1080
DB 1021 AGAGACCGATTAAACAAAACAGAGTTCAGACCCCTTTTCTTTTGGGTGTATCACTAGCT 1080

QY 1081 GCAGGTGCTGTGTTCTTGTAGTGTATCTATTTGACTTATACAGTTTACAGTTGACCATGG 1140
DB 1081 GCAGGTGCTGTGTTCTTGTAGTGTATCTATTTGACTTATACAGTTTACAGTTGACCATGG 1140

QY 1141 AGTGGCAGGTTTTATTCATTGTGGGATCTGGGTATGCAAAATAACACATTCCAATTATT 1200
DB 1141 AGTGGCAGGTTTTATTCATTGTGGGATCTGGGTATGCAAAATAACACATTCCAATTATT 1200

QY 1201 GCATCAGTGTCTGAGCATCAACCTAGCCTGGGTGTCTTCTTCTTGTATCTACATATT 1260
DB 1201 GCATCAGTGTCTGAGCATCAACCTAGCCTGGGTGTCTTCTTCTTGTATCTACATATT 1260

QY 1261 CTTGTATGTACCTTCCAGCAGGCTTTGGTCTGCTCATCAAAATATCAACGATGAAAGA 1320
DB 1261 CTTGTATGTACCTTCCAGCAGGCTTTGGTCTGCTCATCAAAATATCAACGATGAAAGA 1320

QY 1321 GTATTTTGTGCTTATATGCAATCAGTGTCTACTTGTGAGTGTATGCTGCGACTG 1380
DB 1321 GTATTTTGTGCTTATATGCAATCAGTGTCTACTTGTGAGTGTATGCTGCGACTG 1380

QY 1381 ATGTTGACTTTGACTCGTGTGTATGTCTGCTGCAATTTGCTTTTCAAAATGTTTTT 1440
DB 1381 ATGTTGACTTTGACTCGTGTGTATGTCTGCTGCAATTTGCTTTTCAAAATGTTTTT 1440

QY 1441 GAGCCTATTTGGGGATGACATGAAAGGGAATCCACCTGTGGAGGACAGCATGAT 1500
DB 1441 GAGCCTATTTGGGGATGACATGAAAGGGAATCCACCTGTGGAGGACAGCATGAT 1500

QY 1501 GAGGATGACAAAAGAAACCAAGGAAATTTGTATGATAGGCAAGTAAAGTGGGAAACAT 1560
DB 1501 GAGGATGACAAAAGAAACCAAGGAAATTTGTATGATAGGCAAGTAAAGTGGGAAACAT 1560

QY 1561 GCACTGACAGGAAAGAACTGAAGGAGTATAGGCTTAAATATAAAGAGTGTCTACC 1620
DB 1561 GCACTGACAGGAAAGAACTGAAGGAGTATAGGCTTAAATATAAAGAGTGTCTACC 1620

QY 1621 ATGTTGATGTGATGTATTTGATGTATTTGCTGCTCACTGTACCTGGGTCAACAGCAAT 1680
DB 1621 ATGTTGATGTGATGTATTTGATGTATTTGCTGCTCACTGTACCTGGGTCAACAGCAAT 1680

QY 1681 GCCTACTGTAGTCCAAAGTGTAGTCTGCTCATCAATCATGATGGCACCAGGAAATATC 1740
DB 1681 GCCTACTGTAGTCCAAAGTGTAGTCTGCTCATCAATCATGATGGCACCAGGAAATATC 1740

QY 1741 TTAGATGATTTTAGAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAAGA 1800
DB 1741 TTAGATGATTTTAGAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAAGA 1800

QY 1801 GTATGCTCTGGGTGATTTATGCTATCAGATAGTGGATGCTTAATAGAACTACGTTG 1860
DB 1801 GTATGCTCTGGGTGATTTATGCTATCAGATAGTGGATGCTTAATAGAACTACGTTG 1860

QY 1861 GTGATTAATAACACCTTGAATTAACCCACATAGCCTGTGGTGGGAAAGTATGCTTCT 1920
DB 1861 GTGATTAATAACACCTTGAATTAACCCACATAGCCTGTGGTGGGAAAGTATGCTTCT 1920

QY 1921 AATGAACACAGCAGCTTAAATCATAGGACTCTAGATGTAGATATGTTTGGTATT 1980
DB 1921 AATGAACACAGCAGCTTAAATCATAGGACTCTAGATGTAGATATGTTTGGTATT 1980

QY 1981 TTTGGAGGGGTATTGGCTATTCTGGTGTATATCAACAAATTTCTCTGGATGTTAGG 2040
DB 1981 TTTGGAGGGGTATTGGCTATTCTGGTGTATATCAACAAATTTCTCTGGATGTTAGG 2040

QY 2041 ATAGCTGAAGGAGAAATCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGGA 2100
DB 2041 ATAGCTGAAGGAGAAATCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGGA 2100

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 22:52:10 ; Search time 10321.1 Seconds
(without alignments)
11367.538 Million cell updates/sec

Title: US-10-028-384-1

Perfect score: 2481

Sequence: 1 atggcgagagccctcgcccc.....tatctaagaagactgttttaa 2481

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 2364484745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_pa.*

2: gb_hlg.*

3: gb_in.*

4: gb_cm.*

5: gb_ov.*

6: gb_pet.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2481	100.0	2481	6	AX799082 Sequence
2	2481	100.0	2481	9	AY074880 Homo sapi
3	2081.8	83.9	4236	10	BC052433 Mus muscu
4	2078.6	83.8	2710	6	AX799084 Sequence
5	1328.8	53.6	1664	6	CQ782554 Sequence
6	1328.8	53.6	1664	6	BD127193 Sequence
7	1328.8	53.6	1664	9	AK074587 Homo sapi
8	1307.4	52.7	1734	10	BC003206 Mus muscu
9	972	39.2	2546	6	AX099510 Sequence
10	972	39.2	2546	3	AF132552 Drosophil
11	955.8	38.5	2417	6	BD063986 Secreted
12	955.8	38.5	2417	6	AX799088 Sequence
13	955.8	38.5	2699	6	CQ589353 Sequence
14	954	38.5	2510	6	AX136135 Sequence
15	954	38.5	2510	6	BD123520 Homo sapi
16	954	38.5	2510	9	AK075380 Homo sapi
17	923.2	37.2	2503	6	AX017997 Sequence
18	923.2	37.2	2503	6	BD137303 Human nuc
19	861.4	34.7	2284	6	AX882932 Sequence

20	861.4	34.7	2284	6	BD160013	BD160013 Primer fo
21	861.4	34.7	2284	9	AK027789	AK027789 Homo sapi
22	833.4	33.6	2520	5	CR386955	CR386955 Gallus ga
23	761.6	30.7	4522	6	CQ589352	CQ589352 Sequence
24	761.6	30.7	162921	3	AC007853	AC007853 Drosophil
25	761.6	30.7	181132	3	AC008206	AC008206 Drosophil
26	761.6	30.7	227219	3	AE003750	AE003750 Drosophil
27	754.4	30.4	957	9	BC015880	BC015880 Homo sapi
28	753.2	30.4	2508	10	BC013054	BC013054 Mus muscu
29	749.6	30.2	75550	2	AC018145	AC018145 Drosophil
30	712.8	28.7	787	6	AX869456	AX869456 Sequence
31	712.8	28.7	787	6	BD149518	BD149518 Primer fo
32	695.8	28.0	2481	5	BC067313	BC067313 Xenopus t
33	689.2	27.8	22566	6	AX489023	AX489023 Lycopersi
34	688.8	27.8	2555	6	BT012881	BT012881 Sequence
35	671	27.0	764	6	AX136480	AX136480 Sequence
36	671	27.0	764	6	BD123720	BD123720 Secretary
37	670.6	27.0	2736	10	BC037612	BC037612 Mus muscu
38	670.6	27.0	3094	6	AX799090	AX799090 Sequence
39	670.6	27.0	3094	10	MUSTRPR	L34260 Mus musculu
40	667.4	26.9	3666	10	AX129027	AX129027 Mus muscu
41	655	26.4	2855	6	CQ573675	CQ573675 Sequence
42	646.6	26.1	2118	9	BT007100	BT007100 Homo sapi
43	646.6	26.1	2118	12	BT008132	BT008132 Synthetic
44	646.6	26.1	2293	9	HUMTMC	L47337 Homo sapien
45	646.6	26.1	2450	9	BC048348	BC048348 Homo sapi

ALIGNMENTS

RESULT 1
AX799082
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES

AX799082
Sequence 1 from Patent WO03054008.
AX799082
AX799082.1 GI:37605057
Homo sapiens (human)
Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
Perreault,C. and McBride,K.
Mammalian SIMP protein, gene sequence and uses thereof in cancer
therapy
Patent: WO 03054008-A 1 03-JUL-2003;
Compatisgene Inc. (CA)

Location/Qualifiers
1..2481
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..2481
/codon_start=1
/product="SIMP"
/protein_id="CAE48447.1"
/db_xref="GI:37605054"
/translation="MAEPAESKHKSLSNSPWSGLMALGNSRHGHGPGACAHKA
AGGAAPPAPAGLSGSLSPAGWQSLSTFLFLAWAGFSRLFAVIRFESITIEHF
DFWNYRTHHLASHGFYFLNWFDERAWPLRVGGTVYVGLMTHLWILNTL
NITHIRVVCVFLAPTFESGLTSISTFLTRLMNQAGLLAACFIAIVPGYISRSVAG
SPDNIGIAIFALQTYLWKSQVTFVMTCCLSYFYVWAGWGYVFIINLPLH
FVLLLMQRYKRVYIAYSTYIVGLILSQIPIVGFQPIRTSEHMAAGVFLALQAY
AFVLLQRLTKQSFOTLFFGLVSLAAGVFLSVILTYTYGTAPWAGSYVSLWDYGY
AKIHPIIASVSEHQPOTWSPFFDLHLACTPAGIWFCKININRDERVVALYISA
VYFAGVWFLMLTLTPVVCMLSAIAFNVEHILGDDMKRENPFVSDSDKRNQG
NLYDAGKVRKHTEQEKTEGLGPNKISVITMLMLMLLMFAVCHTWTWTSYSPS
TWNSYHGDGTNLDLDFREAYKIRLQNTDEHARVMSWSDYGYOAGMARITLVNPN
TNNSHIALVGKMSNETAAYKIRLQNTDEHARVMSWSDYGYOAGMARITLVNPN
EGEHPDIRESDYTPQGEFVDRKAGSPTLNCILMYKSYRFGECQQLFRTPPGDR
TINAIGNKDIKFKHLBEAFTSEHVLVRIYKVKAPDNRETLDHKPRVTNIFPKQKLS
KTTTKRGYIKNKLVFKKKKISKKTV"

QY 491 TTCAATAGAGAGCTATGTGTTCTTCTGACCAACTTTTAGCGCCTTACATCTATAT 550
Db 411 TCCATAATCGTGACATCTGGTGTCTCTGGCGCGGATCTTCACTGGCCTGACCTCCATCT 470
QY 551 CTACTTCTCTGCTTACAGAGAGCTTTGGAAACCAAGSAGCAGACTTTTATGCTGCTGTT 610
Db 471 CCACCTACCTTGTGACCAAGAGCTGTGTCTCGCGCGCGGCTCTTTCGCGCCAGCT 530
QY 611 TTATTTGCTATGTACAGAGCTACATATCTGGTCTAGTCTGATCTGATCTTGTATATGAAG 670
Db 531 TCATCGGCATCTGCTGCTGCTACATCAGTAGTCTGGTCTGATCTGATCTGATCTGATCTG 590
QY 671 GCATTTGCTATTTGCACTTCACTTCACTATCTTATTTAGGTGAAATCTGTAAAGCTG 730
Db 591 GCATTTGCTATTTGCACTTCACTTCACTATCTTATTTAGGTGAAATCTGTAAAGCTG 650
QY 731 GGTGCTATTTGCACTTCACTTCACTATCTTATTTAGGTGAAATCTGTAAAGCTG 790
Db 651 GATCGGTGTTCTGGTGGCGGCGAGCGCTTTGCTCTTCTTACATGCTGCTGCTGCTG 710
QY 791 GTGCTTATGCTTATCATCAATCTTATTTAGGTGAAATCTGTAAAGCTG 850
Db 711 GTGCTTATGCTTATCATCAATCTTATTTAGGTGAAATCTGTAAAGCTG 770
QY 851 AGAGATACAGCAAAAGAGCTTACATAGCATATAGCACTTTTACATTTGCTGCTTATAT 910
Db 771 GCAGGTACTCGCGCGCTCTGCTGACAGCTTACAGCACTTCTTACATCTGCTGCTGCT 830
QY 911 TATCAATGCGAGATCTTTTGGGATTCAGCGCATCAGAACAGTGAACACATGGCAG 970
Db 831 TCTCCATGCGAGATCTTTGCGGATTCAGCGCATCAGAACAGTGAACACATGGCAG 890
QY 971 CTGAGGTGCTTTGCACTTCTCAAGCTTATGCTTTCTTTCAGATCTGAGAGCCGAT 1030
Db 891 CGCTGGAGTGTGCTGCTCTTATGCGCGTGGCCACCTTGGCCATTTGCACTGCTGCTG 950
QY 1031 TAACAAACAGAGTCTTCCAGACCTTTCTTTTGGGTATCTACATGCTGAGTGTGCTG 1090
Db 951 TGTGCGGACAGAGTCTTCCGAGCTGTTATGCTGCGCGGATGCTGCTGCGGCTGCTG 1010
QY 1091 TGTTCTTATGCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 1150
Db 1011 TCTTTGTGGCGCTGCTGCTGCTTATGCGCGTGGCCACCTTGGCCATTTGCACTGCTG 1070
QY 1151 TTTTATCTTATGCTGCTGCTGCTTATGCGCGTGGCCACCTTGGCCATTTGCACTGCTG 1210
Db 1071 TCTTATCTGCTGCTGCTGCTGCTTATGCGCGTGGCCACCTTGGCCATTTGCACTGCTG 1130
QY 1211 CTGAGCATCAACCTTACAGCTTGGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1270
Db 1131 CGGAGCATCAACCTTACAGCTTGGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1190
QY 1271 CTTTCCAGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1330
Db 1191 CTTTCCAGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1250
QY 1331 CTCTATATGCAATCAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1390
Db 1251 TGCTGTACGCGCATCAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1310
QY 1391 TGCTGTACGCGCATCAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1450
Db 1311 TCAGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1370
QY 1451 TGGGGGATG-----CATGAAAGGAGAAATCCACCTGTGGAG-----GACAGAGTG 1498
Db 1371 TGCAGAGGATTCGCTTAAGGAAATGGGACAGCCATTAAGCGAGCGACCCGAAAGTGGATG 1430
QY 1499 ATGAGGATGACAAAGAAACCAAGGAAATTTGATGATAAGGAGGAGTAAAGTGAAGAAC 1558
Db 1431 AAGCTGAGGATTCATTTGAGAGAGAGAGCTGTATCGAAGGCTGCGAGCTGAGAGATC 1490

QY 1559 ATGCAACTGACAGGAAATAAAGGAGGATTTAGGCGCTTATATATAAAGCAATTGTCTA 1618
Db 1491 GTACTAAGCATGATCCCGCAGGAGATTTAGGCGCTTATATATAAAGCAATTGTCTA 1550
QY 1619 CCATGTTGATGCTGATGCTATTTGATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1678
Db 1551 TTTTGGCGCTTCTAATGCTGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1610
QY 1679 ATGCTTACTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1735
Db 1611 ATGCTTACTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1670
QY 1736 ATATCTTATGATGCTTATGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATG 1795
Db 1671 ACATTTTATGAGCAATTTAGAGAGCTTACTTACTGCTTTTCGAGAACACTTCCGATG 1730
QY 1796 CACGAGTATGCTTGGTGGGATTTAGGCTTATGAGTCTGAGTCTGAGTCTGAGTCTA 1855
Db 1731 CTGCGCTTATGCTTGGTGGGATTTACGAGTACCGGATAGCGGGAATGGCAAGCAG 1790
QY 1856 CGTTGCTGATGATTAACACCTTGGGATTAACAGCCACATAGCACTGGTGGGAAAGCTATGT 1915
Db 1791 CCGTAGTGGATTAATACCTTGGGATTAACAGCCACATAGTACATAGCGCTGGTGGCA 1850
QY 1916 CTTCTAATGAACAGCAGCTTATTAATCATGAGGACTAGATGATGATATGTTTGG 1975
Db 1851 CTTCAACGAGGAGAGCTTCTTACGAAATTTAGCATCTCTTGGCTGGGACTACGTTTGG 1910
QY 1976 TTATTTTGGAGGGTATTTGCTTATTTCTGCTGATGATATCAACAAATTTCTCTGGATG 2035
Db 1911 TGATCTTTGGCGGTGATCGCTTATTTCTGCGATGATATCAACAGTTCCTCTGGATG 1970
QY 2036 TTAGGATGCTGAGGAGAAATCCCAAGACATTCGGAAGAGTACTATTTTACCCAC 2095
Db 1971 TCCGAATTTCTGAGGAGAGCATCCCAAGGACATTTAAGGAAAGCGATTACTTTTACC 2030
QY 2096 AGGAGAAATTTCTGCTAGACAAAGCAGGATCCCTTCTTGTGAAATGCTTATGTATA 2155
Db 2031 GCGGTGATTTAGGCTGATGCTGCGGAGTCTTCCGCGCTTCTCACTGCTTATGTATA 2090
QY 2156 AATGCTACTACTAGATTTGAGAAATGAGCTGAGTGTGCTGCTGCTGCTGCTGCTGCT 2215
Db 2091 AATTAAGCTACTACTAGATTTGAGAAATGAGCTGAGTGTGCTGCTGCTGCTGCTGCT 2150
QY 2216 ACCGAACTGATTAATGCTGAGATTTGGAATAAGGACATTAATTTCAACATTTGGA 2275
Db 2151 ATCGCACAGTAAACGCGCTCATTTGGGATTAAGGACTTCGATCTGACCTTACTGGAG 2210
QY 2276 CCTTTTATCATGAAACACTTGGCTTGTAGGATATATAAGTAAAGGACCTGATAACAG 2335
Db 2211 CCTACACCAAGAACTGCTGCTTGTTCGATCTTATAGGTTGAAGAGCCGATGAGTTCA 2270
QY 2336 AGACATTTAGATCAAAAACCTCGAGTCAACAACTTTTCCCAAAAACAGAAATTTGTCAA 2395
Db 2271 ATAGACCTACTGAGACCAAGGAGAGAGAGTCTTCTCCAGCAAACTTCTTTCGAG 2330
QY 2396 AGAGACTTACAAAAGAGAGGCTGCTTACTTAAATAAGCTGTTTTTAAGAAAGCA 2455
Db 2331 AGAACTTAAAGCTGCTGAG---GGCTACATACGAAACCGGCTGTTGTTTAAGGAA 2387
QY 2456 AGAAATATCTTAAGAGAG 2474
Db 2388 AACGAACCTTGAATAAATAC 2406

RESULT 15

AAF93772

ID AAF93772 standard; cDNA; 2510 BP.

XX AAF93772;

XX AC AC

XX AC AC

XX DT 23-MAY-2001 (first entry)

XX

QY 1451 TGGGGGATGA-----CATGAAAGGGAATAATCCACCTGTGGAG-----GACACAGTGT 1498
Db 1362 TGCAGAGGATTCGCTCAAGCGAATGGCAGACCCATAAGCGCAGCCACCGAAGTCGATG 1421
QY 1499 ATGAGGATGACAAAGAGAAACCAAGGAATTTGTATGATAAGCGAGGTAAAGTGGAGAAC 1558
Db 1422 AAGCTGAGGATTCATTGAGAGAGAGACGCTGTACGACAGGCTGCGAGCTGAAGCATC 1481
QY 1559 ATGCACTGAACAGAGAAACCTGAAGAGGATTAAGCCCTTAATATAAAGACNTTGTCA 1618
Db 1482 GTACTAAGCATGATGCCAGCAGGATAGTGGCGTCAGCTCCACCTCGAAGAGTATTGTGA 1541
QY 1619 CCATGTTGATGCTGATGCTATTGATGATTTGCTGTCCACTGTACTCGGTCACAAGCA 1678
Db 1542 TTTTGGCCGTTCTAATGCTGTGATGATTTGCTGTCCACTGCAGTGGGTGACAGCA 1601
QY 1679 ATGCTTACTTGAAGTCAAGTGTAGTCTGCGCTCTATACAA---TCATGATGCGACAGGA 1735
Db 1602 ATGCTTACTCAGTCCCTCCATTGTCTTGGCTTCCACACAGCTCAAGATGGATCCGCA 1661
QY 1736 ATATCTTAGATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATG 1795
Db 1662 ACATTTAGAGGATTTAGAGAGGCTTACTTACTGGCTTCGACAGACACTGCGGATGATG 1721
QY 1796 CACAGTAAATGCTTGTGTGGATTAATGGCTATCAGATAGTGGATGGCTATAGACTA 1855
Db 1722 CTCGCTTATGCTTGTGTGGATTAACGATACAGATAGCGGAATGGCAAAACAGAGCA 1781
QY 1856 CGTTGTGGGATTAACACCTCGGAATAACAGCCACATAGCACTGGTGGGAAAAGCTATGT 1915
Db 1782 CGTAGTGGATTAATAACGTGGACAAATAGTACATAGCGCTGGTGGCAAGCAATGT 1841
QY 1916 CTTCTAATGAACAGAGGCTTAATAATCATAGGACTCTAGATGTAGATTAATGTTTTGG 1975
Db 1842 CTTCAACGAGGAGGAGCTCTACGAAATATGACATCTCTGACGTGGACTACGTTTTGG 1901
QY 1976 TTATTTTGGAGGGGTTATTGGCTATTCTGCTGATGATATCAACAAATTTCTCTGGATGG 2035
Db 1902 TGATCTTTGGGGTGTGATCGGCTATTCTGCGCATGATATCAACAGTTCTCTGGATGG 1961
QY 2036 TTAGATAGCTGAAGGAGACATCCCAAGACATTCGGAAGTGAATTTTACCCAC 2095
Db 1962 TCCGAATGCTGAGGAGAGACATCCCAAGGACATTAAGGAAAGCGATTACTTTACCGACC 2021
QY 2096 AGGAGAAATCCGCTGTAGACAAAGCAGGATCCCTTACTTTGTTGAATTCCTTTATGTATA 2155
Db 2022 GCGGTGAATTCAGGTTAGATGCCGAGGTGCTCGGCGCTGCTCACTGCTTATGTACA 2081
QY 2156 AAATGTCTACTACAGATTTGGAGAAATGCACTGATTTTCGTACACCCCGAGTTTGT 2215
Db 2082 AATTAAAGCTACTACAGATTCGGGGAATTTGAAGTTGACTACAGAGTCCATCTGGATATG 2141
QY 2216 ACCGAACACGTAATGCTGAGATTGGAAATTAAGGACATTAATTTCAACATTTTGGAGAG 2275
Db 2142 ATCGCACACGTAACGCGCTATTGGGAATAAGGACTTCGATCTGACCTACCTGGAGAGG 2201
QY 2276 CTTTACATCAGACACTGGCTTTAGGATATATAAGTAATAAGCACTGTATAACAGG 2335
Db 2202 CCTACACACAGACACTGGCTTTGTCGCATCTATAGGGTGAAGAGCGCATGATTTCA 2261
QY 2336 AGACATTAGATACAAACCTCGAGTACCAACATTTTCCCAAAACAGAGATTTTGTCAA 2395
Db 2262 ATAGACCATCACTCAAGACCAAGAGAGACGATTCCTCCAGCAAACTTCAATTCGAGAA 2321
QY 2396 AGAGACTTACAAAGAGAGGCTGGCTTACATTAATAAATAGCTGGTTTTTAAAGAGCA 2455
Db 2322 AGAAGCTTAAAGCTGCGCAAG---GGCTACATACGAACCGGCTGTGTTTAAAGGAA 2378
QY 2456 AGAAATATCTAAGAGAC 2474
Db 2379 AACGAACCTTGAATAAAC 2397

RESULT 14
ABL13247
ID ABL13247 standard; cDNA; 2699 BP.
XX AC ABL13247;
XX AC ABL13247;
XX AC ABL13247;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
DE Drosophila; developmental biology; cell signalling; insecticide;
XX Drosophila; gene; ss.
XX Drosophila melanogaster.
OS Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB69144.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 34223; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signaling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at fip.wipo.int/pub/pub/published_pct_sequences

QY Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other;

Query Match 38.5%; Score 955.8; DB 4; Length 2699;

Best Local Similarity 64.8%; Pred. No. 2.8e-236;

Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;

QY 191 AGCGGCTGGTGGCAGTGGCTTCTCTCTTCCACCATCTCTTCTGCTGCTGGTGGCG 250

Db 111 AGTGGCTGGCTACAGCAGCCTAATCACCTTCCCATCTCTGCTAATCGCTGGTGGCG 170

QY 251 GCTTACGCTGGCGCTTTTCGCGCTCATCGCTTCGAAAGCATCATCCAGGTTGAC 310

Db 171 GATTTCTCTCGCTTTCGCGCTCATCGCTTCGAGTTCGATTCATCCAGTTGATC 230

QY 311 CGTGGTTTAACTATAGATCAACACATCATCTTTCATCTCATGGTTCCTATGAATTTTAA 370

Db 231 CGTGGTTTAACTTACCGGGCCACGGCTACATGGTGCAGATGGTGGTGAACCTTCTCA 290

QY 371 ATTGGTTTGAAGAGCATGGTATCCACTAGGAGAAATAGTAGGTGGTACTTTTACC 430

Db 291 ACTGGTTTCAGAGCGCATGGTATCCGCTCGGAGGATTTGTGGCGGTACCGTCTATC 350

QY 431 CAGGTTGATGATAACCGCTGGCTTATTCATTTGATTTTAAATACATTTGAACATACTG 490

Db 351 CCGGCTGATGATACGTCGCGGGAATCCATTTGCTGCTGACGCTACTCAATACCGG 410

29-JAN-2004 (first entry)
Drosophila melanogaster STT3 gene sequence.
source of immunodominant MHC-associated peptide; SIMP; MHC;
major histocompatibility complex; human leukocyte antigen; HLA;
cytotoxic; immunosuppressive; antisense therapy; gene therapy; cancer;
lung cancer; intestine cancer; sarcoma; prostate cancer;
testicular cancer; breast cancer; melanomas; pancreatic cancer;
haematological cancer; immune response; lymphoid cell proliferation;
autoimmune disease; transplant rejection; SIMP-derived peptide;
fruit fly; gene; ds; STT3.
Drosophila melanogaster.
WO2003054008-A2.
03-JUL-2003.
18-DEC-2002; 2002WO-CA001967.
20-DEC-2001; 2001US-00028384.
(COMP-) COMPATIGENE INC.
Perreault C, McBride K;
WPI; 2003-559122/52.
P-PSDB; ADD94790.
New human source of immunodominant MHC-associated peptide (SIMP) nucleic
acids and proteins, useful for diagnosing and treating cancers, e.g. lung
or breast cancer, or for suppressing an immune response in an autoimmune
disease.
Claim 6; SEQ ID NO 7; 66pp; English.
This invention relates to a novel isolated or purified human protein,
termed source of immunodominant major histocompatibility complex (MHC) -
associated peptide (SIMP), which is expressed ubiquitously in human
cells where the protein has the potential of generating several protein
fragments binding with high affinity to a human leukocyte antigen (HLA)
molecule. The invention may allow development of therapeutics with
cytotoxic or immunosuppressive activity or provide sequences useful for
antisense therapy or gene therapy. The source of immunodominant MHC-
associated peptide (SIMP) nucleic acids, proteins and fragments are
useful for diagnosing and treating cancers, for example lung cancer,
intestine cancer, sarcoma, prostate cancer, testicular cancer, breast
cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
proteins are also useful for modulating an immune response. Decreasing
lymphoid cell proliferation is useful for suppressing an immune response
responsible for an autoimmune disease or a transplant rejection. The
present sequence is that of the Drosophila melanogaster STT3 gene which
is related to the invention. Note: This sequence does not appear in the
specification but was obtained by the indexer from GenBank.
Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;
Query Match 38.5%; Score 955.8; DB 10; Length 2417;
Best Local Similarity 64.8%; Pred. No. 2.6e-236;
Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;
191 AGCGGGTGGTGGGCGAGTCGCTCTCTCTTACCATCTCTCTCTGGGCTGGGCTGGCG 250
102 AGTGGCTGGCTACAGAGCGCTAATCACCTTCGCCATCTCTGCTAATCGCTGGCTGGCG 161
251 GCTTCAGCTCGGCTCTTCGGCGTCATCGCTTCGAAAGCATCATCACAGTTCGACC 310
162 GAATTTCCCTCGCTCTTCGGCGTCATCGCTTCGAGTCGATTATCCATGATGTTGATC 221
311 CGTGGTTAATCTAGATCAACATCATCTTCATCTCATGGTTCATGATTTTAA 370
222 CGTGGTTCACTACCGGCGCCACCGCTACATGCTGCGAGATGGTTCGTACATCTCTCA 281

QY 371 ATTGTTTTCATGAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGGTACTGTTTACC 430
DB 282 ACTGGTTTCACGAGCGCGCATGGTATCCGCTCGGAGGATTTGGGGGGGTACCGTCTATC 341
QY 431 CAGGGTTGATGATAACCGCTGGCCCTTATTCATTGGATTTTAAATACATTTGAACATACTG 490
DB 342 CCGGCTGATGATTACGTCGCGGGAATCCATTCGCTGCTGCACGCTACTCAACATACCGG 401
QY 491 TTCACATAGAGACGCTATGTGTTCCTTGCACCAACTTTTAGCGGCTTACATCTATAT 550
DB 402 TCCATATTCGTGACATCTGGGTTCTTGGCGCGCATCTTCAGTGGGCTGACCTCCATCT 461
QY 551 CTACTTTCCTGTTTACAAGAGAACTTTTGGAAACCAAGGAGCAGGACTTTTAGCTGCTTGT 610
DB 462 CCACCTTACCTGCTGACCAAGGAGCTGTGTCTCGGGGCGCGGCTCTTTCGCGCCAGCT 521
QY 611 TTATTGCTATTGTACAGGCTACATATCTCGGTGAGTAGCTGGATCCTTTGATAATGAAG 670
DB 522 TCATCGCCATCGTCTGCTGCTACATCAGTAGGTGGGCTGGATCTGCATACAGG 581
QY 671 GCATTGCTATTTTGCACCTTCAGTTTACATACTATTTTATGGGTAAATCTGTAAAACTG 730
DB 582 GCATTGCCATATTCGCCCTGCAGTTTCACTTCTCTGTGGGTGCGCTCAGTGAAGACTG 641
QY 731 GGTCAAGTTTTTGGACAATGTGCTGCTGCTTATCCTATTCTATATAGTCTCTGCTTGG 790
DB 642 GATCCGTGTTCTGGTGGCGCGCGCTTTGTCTCTTCTACATGGTTCGCGCTGGG 701
QY 791 GTGGTTATGTAATTTATCATCAATCTTATTCACCTGCATGTATTTGTGTTGTACTGATGC 850
DB 702 GTGGCTACGTGTTTCATCATCACTGATACCTCTGCAGCTCTTCTGTACTGCTCATATGG 761
QY 851 AGCATACAGCAAAAGAGCTTACATAGCATATAGCATTTCTACATTTGGGTTTATAT 910
DB 762 GCAGGTACTCGCGCGCTGCTGTGACCACTACAGCACTTCTACATCTGGAGACTGCTGT 821
QY 911 TATCAATGAGATACCTTTTGTGGGATTCAGCAATCAGAAAGTGAACACATGGCAG 970
DB 822 TCTCCATGAGATCCCTTCTGGGGATTTCCAAACGATACGACCACTGAACACATGGCTG 881
QY 971 CTGAGGTGCTTTGCAATGTGCAAGCTTATGCTTCTTGCAGTACTTGCAGACCCGAT 1030
DB 882 CGTGGAGTGTGTTGTGCTCTTATGGCGGTGGCCACCTTGGCCATTTGCAGTCCGTGC 941
QY 1031 TAACAAAAACAAGATTCCAGACCCCTTTCTTTTGGGTGTATCTAGCTGCAGGTGCTG 1090
DB 942 TGTGCGCAACAGATTCCGGAAGCTGTTTCACTGTCGGCGGATTGCTGTGGCGCTGGCG 1001
QY 1091 TGTTCCTAGTGTCTATTTGACTTATACAGTTTACATTCACCATGGAGTGGCAGGT 1150
DB 1002 TCTTTGGCGCTGCTGTGCTCACCATGCTGGGCGTTGTGGCCCTGGGAGTGGAGCT 1061
QY 1151 TTTATTTCATTGTGGGATCTGGGTATGCAAAATACACATTTCCAAATTTATTCATCAGTGT 1210
DB 1062 TCTACTCGCTGGGATCTGCTACGCAAGATCCACATTCCTCATTCATTCCTGCTGT 1121
QY 1211 CTGAGATCAACTACGACTGGGTGCTTTCTTTTGTGATCTACATATCTTGTATGTA 1270
DB 1122 CGAGCATCAGCCCACTTGGTTCTCGTTCTTCTTTGATCTGCACATCTCTGTGTGCG 1181
QY 1271 CTTTCCAGAGCGCTTTGGTTCTGTCATCAAAATATCAACGATGAAGAGTATTTGTTG 1330
DB 1182 CTTTCCAGTGGGAGTGGTACTGTCATCAAGCAGATCAACGACGAGCGGCTTTTCGTGG 1241
QY 1331 CTCATATGCAATCAGTCTGCTACTTTTGTGGAGTGTGCTGGAGTGTGATGTTGACTT 1390
DB 1242 TGTGTACCCATCAGTGGGTTTACTTCTGCTGGTGTGATGTTGGCTTTGATGTTGACCC 1301
QY 1391 TGACTCCAGTCTGTGTATGCTGTGCAATTTGCTTTTCAAAATGCTTTTGTAGCAGCTATT 1450
DB 1302 TCAGCGCGGTGGTGTGATGCTGTGGCGGAGTGGGCTTTTCGGGACTGTGTGATGTTTC 1361

QY 2410 AGAAGCGTGGTACATTAATAAGCTGTTTAAAGAGGCAAGAAATATCTAAG 2469
 DB 901 AGAAGCGTGGTACATTAATAAGCTGTTTAAAGAGGCAAGAAATATCTAAG 960
 QY 2470 AAGACTGTTTAA 2481
 DB 961 AAGACTGTTTAA 972

RESULT 12
 ID AAF98463 standard; cDNA; 2546 BP.

AC AAF98463;
 DT 07-JUN-2001 (first entry)
 XX Human cDNA clone CT585_1 sequence 38Q ID 150.
 XX Human; secreted protein; nutrient; cytokine modulator; proliferation;
 KW differentiation; immune system modulator; tissue growth; chemotactic;
 KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss;
 KW haematopoiesis.

OS Homo sapiens.
 PN WO200119988-A1.
 XX 22-MAR-2001.
 XX 14-SEP-2000; 2000WO-US025135.
 XX 17-SEP-1999; 99US-00398829.
 XX (GEMY) GENETICS INST INC.

PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
 XX WPI; 2001-244801/25.
 DR P-PSDB; AAB90727.
 XX Isolated nucleic acids encoding polypeptides, useful for modulating e.g.
 PT cytokine and cell proliferation/differentiation activity, the immune
 PT system and hematopoiesis regulating activity.
 XX Disclosure; Page 476-477; 557pp; English.

CC Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
 CC proteins AAB90657 - AAB90750. The cDNA clones are isolated from various
 CC tissue types, and may be used in the prevention, treatment and diagnosis
 CC of diseases associated with inappropriate protein expression. The
 CC polypeptides and nucleic acids may be used as nutrients or to modulate
 CC cytokine and cell proliferation/differentiation activity and may also be
 CC involved in modulation of the immune system. The cDNA sequences,
 CC proteins, their agonists and/or antagonists exhibit haematopoiesis
 CC regulating activity; tissue growth activity; activin/inhibin activity;
 CC chemotactic/chemokinetic activity; haemostatic and thrombolytic activity;
 CC receptor/ligand activity; anti-inflammatory activity; haematopoiesis
 CC activity; cadherin/tumour suppressor activity; and/or tumour inhibition
 CC activity. Included in the invention are probes represented in AAF98490 -
 CC AAF98572 which are specific for the cDNA clones encoding the secreted
 CC proteins

XX Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other;

Query Match 39.28; Score 972; DB 5; Length 2546;
 Best Local Similarity 100.0%; Pred. No. 1.7e-240;
 Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1510 AAAAGAACCAAGGAAATTTGTATGATAGCGAGGTAAGTAGGAGAAACATCAACTGAA 1569

DB 1 AAAAGAACCAAGGAAATTTGTATGATAGCGAGGTAAGTAGGAGAAACATCAACTGAA 60
 QY 1570 CAGGAAAAAAGCTGAGAGGATTTAGCCCTAATAATAAAGCAATTTGTCCACCATTTGATG 1629
 DB 61 CAGGAAAAAAGCTGAGAGGATTTAGCCCTAATAATAAAGCAATTTGTCCACCATTTGATG 120
 QY 1630 CTGATCTATTGATGATGTTTCTGTCCACTGTACTGGGTCAAGCAATGCTTACTCT 1689
 DB 121 CTGATCTATTGATGATGTTTCTGTCCACTGTACTGGGTCAAGCAATGCTTACTCT 180
 QY 1690 AGTCCAAAGTGTAGTCTCTGCTCATACATATGATGACCCAGCAATATCTTAGATGAT 1749
 DB 181 AGTCCAAAGTGTAGTCTCTGCTCATACATATGATGACCCAGCAATATCTTAGATGAT 240
 QY 1750 TTTAGAGAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGACGAGTAAATGCT 1809
 DB 241 TTTAGAGAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGACGAGTAAATGCT 300
 QY 1810 TGGTGGATTTAGGCTTATCAGATAGCTGGAATGGCTAATAGAACTACCTTGGTGATAAT 1869
 DB 301 TGGTGGATTTAGGCTTATCAGATAGCTGGAATGGCTAATAGAACTACCTTGGTGATAAT 360
 QY 1870 AACACCTGGAATAACAGCCACATAGCACTGTGGGAAAGCTATGCTTCTAATGAACA 1929
 DB 361 AACACCTGGAATAACAGCCACATAGCACTGTGGGAAAGCTATGCTTCTAATGAACA 420
 QY 1930 GCAGCCTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTTATTTTGGAGGG 1989
 DB 421 GCAGCCTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTTATTTTGGAGGG 480
 QY 1990 GTTATTGGCTATTCTGGTGATGATATCAAAATTTCTCTGGATGTTTAGATGATGCTGAA 2049
 DB 481 GTTATTGGCTATTCTGGTGATGATATCAAAATTTCTCTGGATGTTTAGATGATGCTGAA 540
 QY 2050 GGAGAACATCCCAAGACATTCGGAAAGTACTATTTTACCCACAGGAGAAATCCGT 2109
 DB 541 GGAGAACATCCCAAGACATTCGGAAAGTACTATTTTACCCACAGGAGAAATCCGT 600
 QY 2110 GTAGACAAAGCGAGATCCCTTACTTTTGGTGAATTCGCTTATGTATAAATGTCTACTAC 2169
 DB 601 GTAGACAAAGCGAGATCCCTTACTTTTGGTGAATTCGCTTATGTATAAATGTCTACTAC 660
 QY 2170 AGATTGGGAGAAATGCGATGCTGATTTTGTACACCCCGAGGTTTGTACCGAACAGTAAT 2229
 DB 661 AGATTGGGAGAAATGCGATGCTGATTTTGTACACCCCGAGGTTTGTACCGAACAGTAAT 720
 QY 2230 GCTGAGATTGGAAATAAGGACATTAATTTCAAAACATTTTGGAGAGAGCCTTTTACATGAA 2289
 DB 721 GCTGAGATTGGAAATAAGGACATTAATTTCAAAACATTTTGGAGAGAGCCTTTTACATGAA 780
 QY 2290 CACTGGCTTTTAGGATATATAAAGTAAAGCACTGATAACAGGAGACATTTAGATCAC 2349
 DB 781 CACTGGCTTTTAGGATATATAAAGTAAAGCACTGATAACAGGAGACATTTAGATCAC 840
 QY 2350 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAGAGACTACCAA 2409
 DB 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAGAGACTACCAA 900
 QY 2410 AGAAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAGGCAAGAAATATCTAAG 2469
 DB 901 AGAAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAGGCAAGAAATATCTAAG 960
 QY 2470 AAGACTGTTTAA 2481
 DB 961 AAGACTGTTTAA 972

RESULT 13
 ADD94789
 ID ADD94789 standard; DNA; 2417 BP.
 XX
 XX ADD94789;
 XX

Db 661 TTTTGGGATCCAGCAATCAGAAAGTGAACATGGCAGTGCAGGTGTCTTTC 720
 QY 987 ATTGCTCAAGCTTATCTTCTTCTGAGTATCTGAGAGACCGATTAAACAAAAGAGTT 1046
 Db 721 ATTGCTCAAGCTTATCTTCTTCTGAGTATCTGAGAGACCGATTAAACAAAAGAGTT 780
 QY 1047 CCAGACCCCTTTCTTTTGGGTGTATCAGTATGAGTGTCTGTCTTCTTGTGTGTAT 1106
 Db 781 CCAGACCCCTTTCTTTTGGGTGTATCAGTATGAGTGTCTGTCTTCTTGTGTGTAT 840
 QY 1107 CTATTTGACATATACAGTTTACATTCACCATGGAGTGGCAGTTTATTCATTTGGGA 1166
 Db 841 CTATTTGACATATACAGTTTACATTCACCATGGAGTGGCAGTTTATTCATTTGGGA 900
 QY 1167 TACTGGGTATGCAAAATATACATTCACATTTATTCATGATGATGATGATGATGATGAT 1226
 Db 901 TACTGGGTATGCAAAATATACATTCACATTTATTCATGATGATGATGATGATGATGAT 960
 QY 1227 GACTTGGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1286
 Db 961 GACTTGGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1020
 QY 1287 TTGTTCTGATCAAAATATACATTCACATTTATTCATGATGATGATGATGATGATGATGAT 1329
 Db 1021 TTGTTCTGATCAAAATATACATTCACATTTATTCATGATGATGATGATGATGATGATGAT 1063

RESULT 11

RAV44866
 ID AAV44866 standard; cDNA; 2546 BP.

XX AC AAV44866;

XX DT 21-OCT-1998 (first entry)

XX DE Clone CT585_1 coding sequence.

XX KW Secreted protein; nutritional source; cell proliferation activity;
 KW cell differentiation activity; immune stimulant; tissue growth activator;
 KW haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor;
 KW tumour inhibitor; clone CT585_1; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX CDS 112..972
 XX FT /*tag= a

XX FN WO9825962-A2.

XX PD 18-JUN-1998.

XX PF 12-DEC-1997; 97WO-US023224.

XX PR 13-DEC-1996; 96US-00766263.

XX PR 11-DEC-1997; 97US-00989232.

XX PA (GEM) GENETICS INST INC.

XX PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
 PI Spaulding V, Agostino MU;

XX DR WPI; 1998-362424/31.

XX DR P-PSDB; AAW69247.

XX PT New isolated polynucleotides - obtained from human adult testis, human
 PT adult ovary, human adult brain and human adult heart cDNA libraries.

XX PS Claim 35; Page 79-81; 108pp; English.

XX CC This sequence represents a polynucleotide of the invention, and encodes a
 CC secreted protein. It was isolated from a human adult brain cDNA library,
 CC and is designated clone CT585_1. The DNA sequences and encoded

CC polypeptides can be used as nutritional sources or supplements, or may
 CC exhibit e.g. cytokine and cell proliferation/differentiation activity,
 CC immune stimulating or suppressing activity, haematopoiesis regulating
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC activin/inhibin activity, chemostatic/chemokinetic activity,
 CC cadherin/tumour invasion suppressor activity, tissue growth activity,
 CC tumour inhibition activity or other activities
 XX

SQ Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other;

Query Match 39.2%; Score 972; DB 2; Length 2546;

Best Local Similarity 100.0%; Pred. No. 1.7e-240;

Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1510 AAAAGAAACCAAGGAAATTTGTATGATAGGAGGATTAAGTGAGGAAACATGCAACTGAA 1569

Db 1 AAAAGAAACCAAGGAAATTTGTATGATAGGAGGATTAAGTGAGGAAACATGCAACTGAA 60

QY 1570 CAGGAAAAACCAAGGAGGATTAAGGAGGATTAAGGAGGATTAAGGAGGATTAAGGAGGAT 1629

Db 61 CAGGAAAAACCAAGGAGGATTAAGGAGGATTAAGGAGGATTAAGGAGGATTAAGGAGGAT 120

QY 1630 CTGATGCTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1689

Db 121 CTGATGCTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180

QY 1690 AGTCCAAGTGTAGTCTCTGCTCATACATCATGATGATGATGATGATGATGATGATGATGATGAT 1749

Db 181 AGTCCAAGTGTAGTCTCTGCTCATACATCATGATGATGATGATGATGATGATGATGATGATGAT 240

QY 1750 TTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGCT 1809

Db 241 TTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGCT 300

QY 1810 TGGTGGATTATGCTATCAGATAGCTGGAATGCTAATGAATGCTAATGCTAATGCTAATGCTAAT 1869

Db 301 TGGTGGATTATGCTATCAGATAGCTGGAATGCTAATGAATGCTAATGCTAATGCTAATGCTAAT 360

QY 1870 AACACTGGAATPAAACAGCCACATAGCAGCTGTGGGAAAGCTATGCTTTTAATGAACA 1929

Db 361 AACACTGGAATPAAACAGCCACATAGCAGCTGTGGGAAAGCTATGCTTTTAATGAACA 420

QY 1930 GCAGCCTATAAATCATGAGGACTCTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1989

Db 421 GCAGCCTATAAATCATGAGGACTCTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 480

QY 1990 GTTATTGGCTATTCTGGTCAATCAACAAATTTCTCGATGATGATGATGATGATGATGATGATGAT 2049

Db 481 GTTATTGGCTATTCTGGTCAATCAACAAATTTCTCGATGATGATGATGATGATGATGATGATGAT 540

QY 2050 GGAGAACATCCCAAGACATTCGGGAAAGTCACTATTTTACCCACAGGAGGATTCCT 2109

Db 541 GGAGAACATCCCAAGACATTCGGGAAAGTCACTATTTTACCCACAGGAGGATTCCT 600

QY 2110 GTAGACAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGATATAAATGTCATAC 2169

Db 601 GTAGACAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGATATAAATGTCATAC 660

QY 2170 AGATTTGGAGAAATCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2229

Db 661 AGATTTGGAGAAATCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720

QY 2230 GCTGAGATTGGAATPAAAGGACATTAATTTCAACATTTGGAAGAGGCTTTTACATCAGAA 2289

Db 721 GCTGAGATTGGAATPAAAGGACATTAATTTCAACATTTGGAAGAGGCTTTTACATCAGAA 780

QY 2290 CACTGCTGTGATGATATATAAGTAAAGCCTGATACAGGAGACATGATAC 2349

Db 781 CACTGCTGTGATGATATATAAGTAAAGCCTGATACAGGAGACATGATAC 840

QY 2350 AAACCTCGAGTCACCAACATTTTCCAAAACAGAGTATTTGTCAAAAGAGACTACAAA 2409

Db 841 AAACCTCGAGTCACCAACATTTTCCAAAACAGAGTATTTGTCAAAAGAGACTACAAA 900

ADG78372 standard; cDNA; 1543 BP.
 ADG78372;
 11-MAR-2004 (first entry)
 Human secreted protein cDNA #123.
 Human; secreted protein; gene: ss; immune disorder;
 haematopoietic disorder; hyperproliferative disorder; infectious disease;
 inflammatory disorder; HIV; anaemia; thrombocytopenia; bleeding; stroke;
 myocardial infarction; Addison's disease; rheumatoid arthritis;
 dermatitis; Grave's disease; multiple sclerosis; Glomerulonephritis;
 diabetes; graft-versus-host disease; inflammatory bowel disease; cancer;
 bacterial infection; viral infection; parasitic infection; osteoporosis;
 wound; atherosclerosis; Alzheimer's disease; Parkinson's disease; ulcer;
 food additive; food preservative.
 Homo sapiens.
 US2003211472-A1.
 13-NOV-2003.
 28-MAR-2001; 2001US-00818683.
 05-MAY-1999; 99US-00305736.
 (FENG/) FENG P.
 (RUBEN/) RUBEN S M.
 (ROSE/) ROSEN C A.
 (EBNER/) EBNER R.
 (OLSEN/) OLSEN H S.
 (NIJ/) NI J.
 (WEI/) WEI Y.
 (SOPPET/) SOPPET D R.
 (MOORE/) MOORE P A.
 (KYAW/) KYAW H.
 (LAFLEUR/) LAFLEUR D W.
 (SHI/) SHI Y.
 (JANAT/) JANAT F.
 (ENDRESS/) ENDRESS G A.
 (CARTER/) CARTER K C.
 Feng P, Ruben SM, Rosen CA, Ebner R, Olsen HS, Ni J, Wei Y;
 Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;
 Endress GA, Carter KC;
 WPI; 2004-051575/05.
 P-PSDB; ADG78500.
 New secreted nucleic acid for diagnosing, preventing or treating diseases
 PT associated with aberrant expression or activity of the polypeptide it
 PT encodes, e.g. cancer, human immunodeficiency virus, Parkinson's disease,
 PT or diabetes.
 Claim 1; SEQ ID NO 133; 377bp; English.
 The invention also relates to human secreted proteins and the nucleic
 CC acids encoding them. The proteins and nucleic acids are useful in
 CC diagnosing, preventing, prognosing or treating diseases or disorders
 CC associated with aberrant expression and/or activity of the secreted
 CC proteins, such as immune disorders, haematopoietic disorders,
 CC hyperproliferative disorders, infectious diseases or inflammatory
 CC disorders. In particular, the diseases or disorders are HIV, anaemia,
 CC thrombocytopenia, bleeding, stroke, myocardial infarction, Addison's
 CC disease, rheumatoid arthritis, dermatitis, Grave's disease, multiple
 CC sclerosis, glomerulonephritis, diabetes, graft-versus-host disease,
 CC inflammatory bowel disease, cancer, bacterial infections, viral
 CC infections, parasitic infections, osteoporosis, wounds, atherosclerosis,
 CC Alzheimer's disease, Parkinson's disease or ulcers. The sequences may
 CC also be used as food additives or preservatives, or for modulating
 CC mammalian mental or physical characteristics. The nucleic acids are also

CC used in chromosome mapping, in forensic biology or as molecular weight
 CC markers. This sequence represents cDNA encoding a human secreted protein
 CC of the invention.
 XX
 SQ Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;
 Query Match 42.7%; Score 1060.4; DB 12; Length 1543;
 Best Local Similarity 99.8%; Pred. No. 1.7e-263;
 Matches 1061; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 267 CTTCCCGCTCATCCGCTTCGAAGCATCATCCAGAGTTCGACCCGCGGTTTAACTATAG 326
 DB 1 CTTCCCGCTCATCCGCTTCGAAGCATCATCCAGAGTTCGACCCGCGGTTTAACTATAG 60
 QY 327 ATCAACACATCATCTTGATCTCATGGTCTTATGAATTTTAAATTTGTTTCAAGAAAG 386
 DB 61 ATCAACACATCATCTTGATCTCATGGTCTTATGAATTTTAAATTTGTTTCAAGAAAG 120
 QY 387 AGCATGGTATCCACTAGGAAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATAAC 446
 DB 121 AGCATGGTATCCACTAGGAAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATAAC 180
 QY 447 CGCTGGCTTATTCATTGGATTTTAAATPACATTAACAATACTGTTCCACATAGAGACGT 506
 DB 181 CGCTGGCTTATTCATTGGATTTTAAATPACATTAACAATACTGTTCCACATAGAGACGT 240
 QY 507 ATGTGTGTTCTTGCAACCACTTTTAGCGGCTTACATCTATATCTACTTCTGCTTAC 566
 DB 241 ATGTGTGTTCTTGCAACCACTTTTAGCGGCTTACATCTATATCTACTTCTGCTTAC 300
 QY 567 AAGAGAACTTTGGAAACCAAGAGGAGGAGCTTTTAGCTGCTGTTTATTCTATTGTACC 626
 DB 301 AAGAGAACTTTGGAAACCAAGAGGAGGAGCTTTTAGCTGCTGTTTATTCTATTGTACC 360
 QY 627 AGCTACATATCTCGGTGAGTGGTCTGTTGATAATGAGGCACTTCTATTGTTTC 686
 DB 361 AGCTACATATCTCGGTGAGTGGTCTGTTGATAATGAGGCACTTCTATTGTTTC 420
 QY 687 ACTTCAGTTCACATATACTATTATGGGTAAATCTGTAAAACTGGGFCAGTTTTTGGAC 746
 DB 421 ACTTCAGTTCACATATACTATTATGGGTAAATCTGTAAAACTGGGFCAGTTTTTGGAC 480
 QY 747 ATGTGCTGCTGCTTATCTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 806
 DB 481 ATGTGCTGCTGCTTATCTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 807 CATCAATCTTATTCACATGCTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 866
 DB 541 CATCAATCTTATTCACATGCTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 600
 QY 867 AGCTACATAGCATATAGCACTTTCTACATGCTGGTAAATATATATATATATATATAT 926
 DB 601 AGCTACATAGCATATAGCACTTTCTACATGCTGGTAAATATATATATATATATATAT 660
 QY 927 TTTTGTGGGATTCACAGCAATCAGAACAGTGAACACATGGCAGCTGCAGGTGCTTTTC 986
 DB 661 TTTTGTGGGATTCACAGCAATCAGAACAGTGAACACATGGCAGCTGCAGGTGCTTTTC 720
 QY 987 ATGTGCTGAGCTTATGCTTTCTTTCAGTATCTGAGAGACGATTAACAAAAACAAGATT 1046
 DB 721 ATGTGCTGAGCTTATGCTTTCTTTCAGTATCTGAGAGACGATTAACAAAAACAAGATT 780
 QY 1047 CCAGACCCCTTTTCTTTTGGGTGATCATCTAGCTGCGAGGTGCTGTGTTCTTATGTGTCAT 1106
 DB 781 CCAGACCCCTTTTCTTTTGGGTGATCATCTAGCTGCGAGGTGCTGTGTTCTTATGTGTCAT 840
 QY 1107 CTATTGTGCTTATACAGGTTTACATGGACCACTGGAGTGGCAGGTTTATTTCATTGTGGGA 1166
 DB 841 CTATTGTGCTTATACAGGTTTACATGGACCACTGGAGTGGCAGGTTTATTTCATTGTGGGA 900
 QY 1167 TACTGGGTATGCAAAAAATACACATTCCCAATTTATGTCATGAGTCTCAGCATCAACCTAC 1226
 DB 901 TACTGGGTATGCAAAAAATACACATTCCCAATTTATGTCATGAGTCTCAGCATCAACCTAC 960

XX PF 04-NOV-1998; 98WO-US023435.
 XX PR 07-NOV-1997; 97US-0064900P.
 XX PR 07-NOV-1997; 97US-0064908P.
 XX PR 07-NOV-1997; 97US-0064911P.
 XX PR 07-NOV-1997; 97US-0064912P.
 XX PR 07-NOV-1997; 97US-0064983P.
 XX PR 07-NOV-1997; 97US-0064984P.
 XX PR 07-NOV-1997; 97US-0064985P.
 XX PR 07-NOV-1997; 97US-0064987P.
 XX PR 07-NOV-1997; 97US-0064988P.
 XX PR 07-NOV-1997; 97US-0066089P.
 XX PR 17-NOV-1997; 97US-0066090P.
 XX PR 17-NOV-1997; 97US-0066094P.
 XX PR 17-NOV-1997; 97US-0066095P.
 XX PR 17-NOV-1997; 97US-0066100P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX PI Feng P, Carter KC, Endress GA, Rosen CA, Ruben SM, Janat F;
 XX PI Ni J, Wei Y, Moore PA, Soppet DR, Kyaw H, Lafleur DW, Olsen HS;
 XX PI Shi Y, Ebner R;
 XX WPI: 1999-337740/28.
 XX P-PSDB: AAY27689, AAY27918, AAY27919, AAY27920, AAY27921, AAY27922,
 XX AAY27923.
 XX New human secreted proteins and coding sequences useful for treating
 XX disorders of the immune system and hyperproliferative disorders.
 XX Claim 1; Page 347; 50pp; English.
 XX This sequence represents a nucleic acid molecule which encodes a secreted
 XX human protein. The gene number is given in the descriptor line. The gene
 XX can be used to generate fusion proteins by linking to the gene to a human
 XX immunoglobulin Fc portion (e.g. AAX84924) for increasing the stability of
 XX the fused protein as compared to the human protein only. The invention
 XX relates to 125 novel genes and their fragments (nucleic acid sequences;
 XX AAX84933-X85057; amino acid sequences AAY27567-Y27933) which are useful
 XX for preventing, treating or ameliorating medical conditions e.g. by
 XX protein or gene therapy. Also, pathological conditions can be diagnosed
 XX by determining the amount of the new polypeptides in a sample or by
 XX determining the presence of mutations in the new polynucleotides.
 XX Specific uses are described for each of the 125 polynucleotides, based on
 XX which tissues they are most highly expressed in (see AAX84933 for
 XX described uses)
 XX Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;
 Query Match 42.7%; Score 1060.4; DB 2; Length 1543;
 Best Local Similarity 99.8%; Pred. No. 1.7e-263;
 Matches 1061; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 267 CTTCCGCGCTCATCCGCTTCGAAAGCATATCCACGAGTTTCGACCCGCTGGTTAACTATAG 326
 DB 1 CTTCCGCGCTCATCCGCTTCGAAAGCATATCCACGAGTTTCGACCCGCTGGTTAACTATAG 60
 QY 327 ATCAACACATCATCTTCATCTCATGCGTTCTAAGAAATTTTAAATGGTTTGAAGAAAG 386
 DB 61 ATCAACACATCATCTTCATCTCATGCGTTCTAAGAAATTTTAAATGGTTTGAAGAAAG 120
 QY 387 AGCATGGTATCCATAGGAAGATAGTAGGTGGTACTGTTTACCCAGGGTTGATGATAAC 446
 DB 121 AGCATGGTATCCATAGGAAGATAGTAGGTGGTACTGTTTACCCAGGGTTGATGATAAC 180
 QY 447 CGCTGGCGCTTATTCATGGATTTTAAATACATTCACATCAACTGTTTCATGAAGAGGT 506
 DB 181 CGCTGGCGCTTATTCATGGATTTTAAATACATTCACATCAACTGTTTCATGAAGAGGT 240
 QY 507 ATGTGTGTCTTCGACCAACTTTTAGCGGCCCTTACATCTATATCTACTTTCCTGCTTAC 566
 DB 241 ATGTGTGTCTTCGACCAACTTTTAGCGGCCCTTACATCTATATCTACTTTCCTGCTTAC 300

QY 567 AAGAGNACTTTGGAACCCAGGAGGAGGACTTTTAGCTGCTTGTGTTTATTCCTATTGTACC 626
 DB 301 AAGAGAACTTTGGAACCCAGGAGGAGGAGGACTTTTAGCTGCTTGTGTTTATTCCTATTGTACC 360
 QY 627 AGGCTACATATCTCGGTTCAGTAGTGGATCTCTTTGATAATGAAGGCAATTCCTATTGTC 686
 DB 361 AGGCTACATATCTCGGTTCAGTAGTGGATCTCTTTGATAATGAAGGCAATTCCTATTGTC 420
 QY 687 ACTTCAGTTTCACATATCTTTATGGGTAAATCTGTAAACTGGGTGAGTTCAGTTTTCGAC 746
 DB 421 ACTTCAGTTTCACATATCTTTATGGGTAAATCTGTAAACTGGGTGAGTTCAGTTTTCGAC 480
 QY 747 AATGTGCTGCTCTTATCTCTATTTCTATATGCTCTCTGCTTGGGTGGTATGATTTAT 806
 DB 481 AATGTGCTGCTCTTATCTCTATTTCTATATGCTCTCTGCTTGGGTGGTATGATTTAT 540
 QY 807 CATCAATCTTATCCACTGCGATGATTTGCTGTTGTTTCTGATGAGAGATACAGCAAAAG 866
 DB 541 CATCAATCTTATCCACTGCGATGATTTGCTGTTGTTTCTGATGAGAGATACAGCAAAAG 600
 QY 867 AGTCTACATAGCATATAGCAGTCTTCTACATTTGCGGTTTAAATATTAATCAATGAGATACC 926
 DB 601 AGTCTACATAGCATATAGCAGTCTTCTACATTTGCGGTTTAAATATTAATCAATGAGATACC 660
 QY 927 TTTTGGGATTCAGCCCAATCAGCAAGTGAACACATGCGAGCTGCGAGTGTCTTTGC 986
 DB 661 TTTTGGGATTCAGCCCAATCAGCAAGTGAACACATGCGAGCTGCGAGTGTCTTTGC 720
 QY 987 ATTGCTGCAAGCTTATGCTTTTTCAGTATCTCAGAGACCGATTAACAAAACAAGATT 1046
 DB 721 ATTGCTGCAAGCTTATGCTTTTTCAGTATCTCAGAGACCGATTAACAAAACAAGATT 780
 QY 1047 CCAGACCCCTTTCTTTTGGGTGATCAGTACCTGCTGCTGCTGCTTCTTGTAGTGTAT 1106
 DB 781 CCAGACCCCTTTCTTTTGGGTGATCAGTACCTGCTGCTGCTGCTTCTTGTAGTGTAT 840
 QY 1107 CTATTTGACTTATACAGGTTACATTTGCCACCATGAGTGGCAGGTTTATTCATTTGGGA 1166
 DB 841 CTATTTGACTTATACAGGTTACATTTGCCACCATGAGTGGCAGGTTTATTCATTTGGGA 900
 QY 1167 TACTGGGTATGCAAAATACATCTCCAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1226
 DB 901 TACTGGGTATGCAAAATACATCTCCAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 QY 1227 GACTTGGGTGCTTTCTTTCTTTGATCTACATATTTTGTATGATGATGATGATGATGATGAT 1286
 DB 961 GACTTGGGTGCTTTCTTTCTTTGATCTACATATTTCTGATGATGATGATGATGATGATGAT 1020
 QY 1287 TTGGTTCGATCAAAATATCAACGATGAGAGATTTTGT 1329
 DB 1021 TTGGTTCGATCAAAATATCAACGATGAGAGATTTTGT 1063

RESULT 8
 ACD18981
 ID ACD18981 standard; cDNA; 1543 BP.
 XX ACD18981;
 XX XX
 DT 21-AUG-2003 (first entry)
 XX XX
 DE Novel human secreted protein cDNA #123.
 XX XX
 KW Human; secreted protein; cytostatic; neuroprotective; hepatotropic;
 KW gene therapy; cancer; liver disorder; hepatitis; neural disorder;
 KW Alzheimer's disease; gene; ss.
 XX OS Homo sapiens.
 XX XX
 PN US2003028003-A1.
 XX XX
 PD 06-FEB-2003.

OS Homo sapiens.
 PN WO200190304-A2.
 XX 29-NOV-2001.
 PD 18-MAY-2001; 2001WO-US016450.
 PF 19-MAY-2000; 2000US-0205515P.
 PR (HUMA-) HUMAN GENOME SCI INC.
 XX PA Birse CE, Rosen CA;
 PI WPI; 2002-122018/16.
 DR P-PSDB; ABB99441.
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders.
 PT Claim 4; SEQ ID NO 412; 2081pp + Sequence Listing; English.
 XX The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB9040-ABB9044) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX Query Match 42.8%; Score 1061; DB 6; Length 1209;
 Best Local Similarity 99.8%; Pred. No. 1.1e-263;
 Matches 1061; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 257 CTTCCCGTCATCCGCTTCGAAAGCATCCACGAGTTCGACCCCGTGGTTAACTATAG 326
 1 CTTCCCGTCATCCGCTTCGAAAGCATCCACGAGTTCGACCCCGTGGTTAACTATAG 60
 327 ATCAACATCATCTTCATCTCATGGTTCTATGAATTTTAAATGGTTTGTATGAAG 386
 61 ATCAACATCATCTTCATCTCATGGTTCTATGAATTTTAAATGGTTTGTATGAAG 120
 387 AGCATGGTATCCATAGGAAGATAGTAGTGTGTCTGTTTACCAGGTTGATGATAC 446
 121 AGCATGGTATCCATAGGAAGATAGTAGTGTGTCTGTTTACCAGGTTGATGATAC 180
 447 CGTGGCTTTATTCATGGATTTTAAATACATGAACATCACTGTTCAATGAAGAGCT 506
 181 CGTGGCTTTATTCATGGATTTTAAATACATGAACATCACTGTTCAATGAAGAGCT 240
 507 ATGTGTGTCTTCACCAACTTTAGCGGCTTATACATCTATATCTACCTTCCTGTTAC 566
 241 ATGTGTGTCTTCACCAACTTTAGCGGCTTATACATCTATATCTACCTTCCTGTTAC 300
 567 AAGGAACCTTGGAAACCAAGGAGGAGCATTTTAGCTGCTGTTTATGCTATTGTACC 626
 301 AAGGAACCTTGGAAACCAAGGAGGAGCATTTTAGCTGCTGTTTATGCTATTGTACC 360

QY 627 AGCTACATATCTCGGTGAGTGTGATCCCTTTGATATGAGGCATTTGCTATTTTGC 686
 DB 361 AGGCTACATATCTCGGTGAGTGTGATCCCTTTGATATGAGGCATTTGCTATTTTGC 420
 QY 687 ACTTCAGTTTCACTACTATTTATGGTAAATCTGTAATACTGGTTCAGTTTTTGGAC 746
 DB 421 ACTTCAGTTTCACTACTATTTATGGTAAATCTGTAATACTGGTTCAGTTTTTGGAC 480
 QY 747 AATGTGCTGCTCTTATCTCTATTTCTATATGCTCTCTCTTTGGGGTGGTTATGATTAT 806
 DB 481 AATGTGCTGCTCTTATCTCTATTTCTATATGCTCTCTCTTTGGGGTGGTTATGATTAT 540
 QY 807 CATCAATCTTATCCACTGATGATTTGTTGTTTACTGATGAGAGATACAGCAAAAG 866
 DB 541 CATCAATCTTATCCACTGATGATTTGTTGTTTACTGATGAGAGATACAGCAAAAG 600
 QY 867 AGTCTACATAGCATATAGCACTTTCTACATTTGTTGGTTTAAATATATCAATGCGATACC 926
 DB 601 AGTCTACATAGCATATAGCACTTTCTACATTTGTTGGTTTAAATATATCAATGCGATACC 660
 QY 927 TTTTGTGGGATTCAGCCCAATCAGAACAGAGTGAACACATGCGAGCTGCGAGTGTCTTTC 986
 DB 661 TTTTGTGGGATTCAGCCCAATCAGAACAGAGTGAACACATGCGAGCTGCGAGTGTCTTTC 720
 QY 987 ATTGCTGCAAGCTTATGCTTTCTTCAGTATCTGAGAGACCGATTAACAAAACAGAGTT 1046
 DB 721 ATTGCTGCAAGCTTATGCTTTCTTCAGTATCTGAGAGACCGATTAACAAAACAGAGTT 780
 QY 1047 CCAGAGCCCTTTCTTTTGGGGTATCACTAGCTGCGAGTGTCTGTTTCTTGTAGTGCAT 1106
 DB 781 CCAGAGCCCTTTCTTTTGGGGTATCACTAGCTGCGAGTGTCTGTTTCTTGTAGTGCAT 840
 QY 1107 CTATTTGACTTATACAGGTTTACATTCACCAATGAGTGGAGTGGAGTTTATCATTTGTTGGA 1166
 DB 841 CTATTTGACTTATACAGGTTTACATTCACCAATGAGTGGAGTGGAGTTTATCATTTGTTGGA 900
 QY 1167 TACTGGGTATGAAATAACATATCCAAATTTCCATGTCAGTGTCTGAGCATCACTTAC 1226
 DB 901 TACTGGGTATGAAATAACATATCCAAATTTCCATGTCAGTGTCTGAGCATCACTTAC 960
 QY 1227 GACTTGGGTGTCTTTCTTTTGTATCTACATATTTCTTGTATGATCTTCCAGCAGGCT 1286
 DB 961 GACTTGGGTGTCTTTCTTTTGTATCTACATATTTCTTGTATGATCTTCCAGCAGGCT 1020
 QY 1287 TTGTTTCTGCAATCAAAATATCAAGCATGAAGAGTATTTGTT 1329
 DB 1021 TTGTTTCTGCAATCAAAATATCAAGCATGAAGAGTATTTGTT 1063
 RESULT 7
 AAX85055
 ID AAX85055 standard; DNA; 1543 BP.
 XX AC AAX85055;
 XX DT 30-JUL-1999 (first entry)
 XX DE Human secreted protein gene No. 123.
 XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 OS Homo sapiens.
 XX WO9924836-A1.
 XX 20-MAY-1999.
 PD

XX DT 20-MAY-2004 (first entry)
 XX DE Full length human cDNA clone SeqID 2694.
 XX KW human; medicine; signal transduction; glycoprotein; transcription;
 XX KW oligo-capping method; ss; gene.
 XX OS Homo sapiens.
 XX PN EP1396543-A2.
 XX PD 10-MAR-2004.
 XX PF 07-JUL-2000; 2003EP-00025638.
 XX PR 08-JUL-1999; 99JP-00194486.
 XX PR 11-JAN-2000; 2000JP-00118774.
 XX PR 02-MAY-2000; 2000JP-00183865.
 XX PR 07-JUL-2000; 2000EP-00114089.
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI; 2004-204755/20.
 XX P-PSDB; ADL30662.
 XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
 XX length human cDNAs.
 XX Example 1; SEQ ID NO 2694; 1340pp; English.
 XX This invention relates to a novel primers useful for synthesizing full
 XX length cDNA molecules that encode human proteins. Specifically, it refers
 XX to secretory or membrane proteins that are potential therapeutic agents/
 XX target molecules in the field of medicine, and in particular genes
 XX encoding proteins that are associated with signal transduction,
 XX glycoproteins and transcription. The present invention describes a method
 XX for efficiently cloning a full length human cDNA from both the 5' and 3',
 XX ends using the oligo-capping method. This polynucleotide sequence is a
 XX full length human cDNA clone of the invention.
 XX SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;
 Query Match 53.6%; Score 1328.8; DB 12; Length 1664;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATGGCGAGCCCTCGGCGCGGAGAGCAAGCAAGTCTCCCTCAACTCGTCCCGGNG 60
 DB 28 ATGGCGAGCCCTCGGCGCGGAGAGCAAGCAAGTCTCCCTCAACTCGTCCCGGNG 87
 QY 61 AGTGGCTCATAGGCGCGGAGAGCAAGCAAGTCTCCCTCAACTCGTCCCGGNG 120
 DB 88 AGTGGCTCATAGGCGCGGAGAGCAAGCAAGTCTCCCTCAACTCGTCCCGGNG 147
 QY 121 GCGCACAGCGCGGCGGAGAGCAAGCAAGTCTCCCTCAACTCGTCCCGGNG 180
 DB 148 GCGCACAGCGCGGCGGAGAGCAAGCAAGTCTCCCTCAACTCGTCCCGGNG 207
 QY 181 GGGCTGTGCGAGCGGCGGAGAGCAAGCAAGTCTCCCTCAACTCGTCCCGGNG 240
 DB 208 GGGCTGTGCGAGCGGCGGAGAGCAAGCAAGTCTCCCTCAACTCGTCCCGGNG 267
 QY 241 TGGCTTCCCGGCTTACGCTCGGCGGAGAGCAAGCAAGTCTCCCTCAACTCGTCCCGGNG 300
 DB 268 TGGCTTCCCGGCTTACGCTCGGCGGAGAGCAAGCAAGTCTCCCTCAACTCGTCCCGGNG 327
 QY 301 GAGTTCGAGCGGCTTAACTATAGATCAACATCATCTTGCATCTCANGGGTCTAT 360
 DB 328 GAGTTCGAGCGGCTTAACTATAGATCAACATCATCTTGCATCTCANGGGTCTAT 387

QY 361 GAATTTTAAATGGTTTTCATGAAAGAGCATGGTATCCACTAGGAAGATAGTAGTGGT 420
 DB 388 GAATTTTAAATGGTTTTCATGAAAGAGCATGGTATCCACTAGGAAGATAGTAGTGGT 447
 QY 421 ACTGTTTACCAGGGTGTGATGATAACCGCTGGCCTTATTCATTGGATTTTAAATACATTG 480
 DB 448 ACTGTTTACCAGGGTGTGATGATAACCGCTGGCCTTATTCATTGGATTTTAAATACATTG 507
 QY 481 AACATAAATCTGTTACATAGAGAGCATGTGTGTGTTCCCTTGACCAACATTTTACGGGCTT 540
 DB 508 AACATAAATCTGTTACATAGAGAGCATGTGTGTGTTCCCTTGACCAACATTTTACGGGCTT 567
 QY 541 ACATCTATATCTACTTCTCTGCTTCAAGAGNACTTGGAAACCAAGAGGAGGACTTTTA 600
 DB 568 ACATCTATATCTACTTCTCTGCTTCAAGAGNACTTGGAAACCAAGAGGAGGACTTTTA 627
 QY 601 GCTGCTGTTGTTTATTTGCTATTGTACAGGCTACATATCTCGGTGAGTGTGATCTCTTT 660
 DB 628 GCTGCTGTTGTTTATTTGCTATTGTACAGGCTACATATCTCGGTGAGTGTGATCTCTTT 687
 QY 661 GATAATGAAGGCAATGCTATTTTTCGACTTCAGTTTCACTACTATTATGGTAAATCT 720
 DB 688 GATAATGAAGGCAATGCTATTTTTCGACTTCAGTTTCACTACTATTATGGTAAATCT 747
 QY 721 GTAAAAAATCGGCTCAGTTTGTGGACAATGTGCTGCTCTTATCTCTATATGATGTC 780
 DB 748 GTAAAAAATCGGCTCAGTTTGTGGACAATGTGCTGCTCTTATCTCTATATGATGTC 807
 QY 781 TCTGCTGGGCTGGTGTATGATTTATCATCATCTTATTCACATGCTGATGTTTGTGTTG 840
 DB 808 TCTGCTGGGCTGGTGTATGATTTATCATCATCTTATTCACATGCTGATGTTTGTGTTG 867
 QY 841 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACCTTCTACATTGTG 900
 DB 868 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACCTTCTACATTGTG 927
 QY 901 GGTTTAATATTATCAATGAGATACCTTTTGGGATTCAGGCAATCAGAACAGTGA 960
 DB 928 GGTTTAATATTATCAATGAGATACCTTTTGGGATTCAGGCAATCAGAACAGTGA 987
 QY 961 CACATGGCAGCTGCGAGGTGCTTTGCAATGCTGCAAGCTTATGCTTCTTCAGTATCTG 1020
 DB 988 CACATGGCAGCTGCGAGGTGCTTTGCAATGCTGCAAGCTTATGCTTCTTCAGTATCTG 1047
 QY 1021 AGAGACCGGATTAACAAACAGAGTTCAGACCCCTTTTCTTTTGGGTTGATCATAGCT 1080
 DB 1048 AGAGACCGGATTAACAAACAGAGTTCAGACCCCTTTTCTTTTGGGTTGATCATAGCT 1107
 QY 1081 GCAGGTGCTGTGTTCTTGTAGTCTCATCTATTGTACTTATACAGGTTACATTGCACCATGG 1140
 DB 1108 GCAGGTGCTGTGTTCTTGTAGTCTCATCTATTGTACTTATACAGGTTACATTGCACCATGG 1167
 QY 1141 AGTGGCAGGTTTATTTATTTGTTGGGATACGTTGGTATGCAAAAATACATTTCCAAATTAT 1200
 DB 1168 AGTGGCAGGTTTATTTATTTGTTGGGATACGTTGGTATGCAAAAATACATTTCCAAATTAT 1227
 QY 1201 GCATCAGTGTCTGAGCATCAACTCAGCTTGGGTTGTTCTTCTTTGTGATCTACATATT 1260
 DB 1228 GCATCAGTGTCTGAGCATCAACTCAGCTTGGGTTGTTCTTCTTTGTGATCTACATATT 1287
 QY 1261 CTTGTATGATCTTCCAGCAGGCTTTGGTTCTGCTCAAAAATATCAACGATGAAGA 1320
 DB 1288 CTTGTATGATCTTCCAGCAGGCTTTGGTTCTGCTCAAAAATATCAACGATGAAGA 1347
 QY 1321 GTATTTGTTGCT 1332
 DB 1348 GTATTTGTTGCT 1359

RESULT 5
 AAH33264
 ID AAH33264 standard; cDNA; 1209 BP.

DE Human full-length cDNA, SEQ ID NO: 2694.
 XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX Homo sapiens.
 OS EP1130094-A2.
 XX 05-SEP-2001.
 XX 07-JUL-2000; 2000EP-00114089.
 XX 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Nishikawa T, Isogai T, Hayaashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI; 2001-524255/58.
 DR P-PSDB; AAM93252.
 XX 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 PT Claim 8; SEQ ID NO 2694; 1380pp + Sequence Listing; English.
 XX The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5' and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesizing the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a full length human cDNA of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in CD-ROM format directly
 CC from EPO
 XX SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;
 Query Match 53.6%; Score 1328.8; DB 4; Length 1664;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 ATGGCGAGCCCTCGGCGCGGAGCAAGCAAGTCGTCCTCAACTCGPCCCGTGG 60
 28 ATGGCGAGCCCTCGGCGCGGAGCAAGCAAGTCGTCCTCAACTCGPCCCGTGG 87
 61 AGTGGCTCATAGCCCTTGGAAACAGCGGCGACGCGCCACCGCGCGGCGCCAGTGC 120
 88 AGTGGCTCATAGCCCTTGGAAACAGCGGCGACGCGCCACCGCGCGGCGCCAGTGC 147
 121 GCGCACAAAGCGGCGGCGCGGCGCGCGCGGCGCGCGCGCGCGCGGCGCGG 180
 148 GCGCACAAAGCGGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGGCGG 207
 181 GGGCTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 208 GGGCTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 267
 241 TGGTTCGCGGCTTCAGTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 300
 268 TGGCTTCGCGGCTTCAGTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 327
 301 GAGTTCGACCGCGGCTTCAGTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCAT 360
 328 GAGTTCGACCGCGGCTTCAGTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCAT 387
 361 GAATTTTAAATGGTTTGGATGAAGAGCATGATGATCCACTAGGAAGATAGTAGGTGGT 420

388 GAATTTTAAATGGTTTGGATGAAGAGCATGATGATCCACTAGGAAGATAGTAGGTGGT 447
 421 ACTGTTTACCAGGCTTGGATGAATACCGCTGCGCTTATTCATGATTTTAAATACATTTG 480
 448 ACTGTTTACCAGGCTTGGATGAATACCGCTGCGCTTATTCATGATTTTAAATACATTTG 507
 481 AACATAACTGTTTCATACAGAGAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 540
 508 AACATAACTGTTTCATACAGAGAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 567
 541 ACATCTATATCTACTTCT 600
 568 ACATCTATATCTACTTCT 627
 601 GCTGCTTGTGTTTATTTGCTATTTGCTATTTGCTATTTGCTATTTGCTATTTGCTATTTGCTATTT 660
 628 GCTGCTTGTGTTTATTTGCTATTTGCTATTTGCTATTTGCTATTTGCTATTTGCTATTTGCTATTT 687
 661 GATATGAGGCAATGCTATTTTGGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 688 GATATGAGGCAATGCTATTTTGGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 747
 721 GTAAATACTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 748 GTAAATACTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 807
 781 TCTGCTTGGGCT 840
 808 TCTGCTTGGGCT 867
 841 TTACTGATGACAGATACAGCAAAAGAGCTCTACATAGCATATAGCATATTTCTTCTCATTTGCTG 900
 868 TTACTGATGACAGATACAGCAAAAGAGCTCTACATAGCATATAGCATATTTCTTCTCATTTGCTG 927
 901 GCTTTAATATATCAATGAGATACCTTTTGGGATTCAGGCAATCCAGCAATCCAGCAATCCAGCAAT 960
 928 GGTTTAATATATCAATGAGATACCTTTTGGGATTCAGGCAATCCAGCAATCCAGCAATCCAGCAAT 987
 961 CACATGGCAGCTGCGAGGTGCTTTGCAATGCTGCAAGCTTATGCTTTCTTTCAGATATCTG 1020
 988 CACATGGCAGCTGCGAGGTGCTTTGCAATGCTGCAAGCTTATGCTTTCTTTCAGATATCTG 1047
 1021 AGAGACCGATTAACAAACAGAGTTCAGACCCCTTTCTTTTGGGCTGATCACTAGCT 1080
 1048 AGAGACCGATTAACAAACAGAGTTCAGACCCCTTTCTTTTGGGCTGATCACTAGCT 1107
 1081 GCAGGTGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1140
 1108 GCAGGTGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1167
 1141 AGTGGCAGGTTTATTCATTTGTTGGATGCTGGTATGCAAAATACACATTCCTCATTTATTT 1200
 1168 AGTGGCAGGTTTATTCATTTGTTGGATGCTGGTATGCAAAATACACATTCCTCATTTATTT 1227
 1201 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1260
 1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1287
 1261 CTTGTATGTACCTTCCGAGCGGCTTGGTCTTGTGATCAAAATATCAAGATGAAGA 1320
 1288 CTTGTATGTACCTTCCGAGCGGCTTGGTCTTGTGATCAAAATATCAAGATGAAGA 1347
 1321 GTATTTTGTGCT 1332
 1348 GTATTTTGTGCT 1359

RESULT 4
 ADL30661
 ID ADL30661 standard; cDNA; 1664 BP.
 XX
 AC ADL30661;

481 AACATAACTGTTCATAGAGAGCGTATGTGTCTCTTGCACCAACTTTTAGCGCCCTT 540
Db AACATAACAGTTCACATAAGAGATGTGTGTATCTCTTGCACCAACTTTTAGCGCCCTT 607
481 ACATCTATCTACTCTTCTGCTTACAGAGAACTTTGGAACCAAGGACGAGCTTTTA 600
Db ACATCCATACTCTGCTTCTGCTTACAGAGAACTTTGGAACCAAGGACGAGCTTTTA 667
601 GCTGCTTGTATTTATGTATTTGACAGGCTACATATCTCGGTGAGTGTGATGCTGCTT 660
Db GCTGCTTGTATTTATGTATTTGACAGGCTACATATCTCGGTGAGTGTGATGCTGCTT 727
661 GATATGAGGAGCTATTTTGTGACCTTCACTTACATATTTTATGAGTGTGATGCTGCTT 720
Db GATATGAGGAGCTATTTTGTGACCTTCACTTACATATTTTATGAGTGTGATGCTGCTT 787
721 GTAAAACTGGGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db GTGAAGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 847
781 TCTGCTTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db TCTGCTTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 907
841 TTACTGATGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTTACATTTG 900
Db CTGCTGATGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTTACATTTG 967
901 GGTATATATATCAATGAGATACCTTTTGTGGGATTCAGCCATACAGCAAGTGA 960
Db GGTATATATATCAATGAGATACCTTTTGTGGGATTCAGCCATACAGCAAGTGA 1027
961 CACATGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db CACATGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1087
1021 AGAGCCGATTAACAAACAGAGTCCAGACCTTTCTTTTGGGTATCAGTGT 1080
Db AGAGCCGATTAACAAACAGAGTCCAGACCTTTCTTTTGGGTATCAGTGT 1147
1081 GCAGGTGCTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db GCAGGTGCTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1207
1141 AGTGGCAGTGTATTTCAATTTGGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db AGTGGCAGTGTATTTCAATTTGGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1267
1201 GCATCAGTGTGCTGAGCATCAACCTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db GCATCAGTGTGCTGAGCATCAACCTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1327
1261 CTTGTATGTACCTTCCAGCAGGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db CTTGTATGTACCTTCCAGCAGGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1387
1321 GTATTTGCTGCTATATGCAATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db GTATTTGCTGCTATATGCAATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1447
1381 ATGTTGATTTGATCTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Db ATGTTGATTTGATCTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1507
1441 GAGCAGTATTTGGGGATGATGATGAAAGGAGAAATCCACCTGTGGAGGACAGAGTGTAT 1500
Db GAGCAGTATTTGGGGATGATGATGAAAGGAGAAATCCACCTGTGGAGGACAGAGTGTAT 1567
1501 GAGATGACAAAGAAACCAAGGAATTTGATGATGAGCAGGTTAAGTGAAGAAAT 1560
Db GAGATGACAAAGAAACCAAGGAATTTGATGATGAGCAGGTTAAGTGAAGAAAT 1627

1561 GCAACTGAACAGAGAAATCTGAGAGGATTTAGCCCTTAATATATATAAAGCAATTTGTACC 1620
Db GTGACAGAGCAAGAGAAATCTGAGAGGATTTAGCCCTTAATATATATAAAGCAATTTGTACC 1687
1621 ATCTTGTATCTGATGCTATTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Db ATCTTGTATCTGATGCTATTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1747
1681 GCCTACTGTAGTCCAAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db GCCTACTGTAGTCCAAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1807
1741 TTAGATGATTTAGAGAGCTTACTTTTGGCTTAAAGCAAAATACAGATGAACATGCAAGCA 1800
Db TTAGATGATTTAGAGAGCTTACTTTTGGCTTAAAGCAAAATACAGATGAACATGCAAGCA 1867
1801 GTATGCTTGTGGGATTTAGCTTATCAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
Db GTATGCTTGTGGGATTTAGCTTATCAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1927
1861 GTGATTAATAACCACTGGAATACAGCCATACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
Db GTGATTAATAACCACTGGAATACAGCCATACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1987
1921 AATGAACAGCAGCTTAAATATCATGAGGATCTGATGATGATGATGATGATGATGATGATGATGAT 1980
Db AATGAACAGCAGCTTAAATATCATGAGGATCTGATGATGATGATGATGATGATGATGATGATGAT 2047
1981 TTTGGAGGCTTATTTGGCTTATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040
Db TTTGGAGGCTTATTTGGCTTATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2107
2041 ATAGTCTGAAGGAGCAATCCCAAGCAATTCGGGAAAGTGAATTTTACCCCAAGGCA 2100
Db ATAGTCTGAAGGAGCAATCCCAAGCAATTCGGGAAAGTGAATTTTACCCCAAGGCA 2167
2101 GAATTCGCTGATGACAAAGCAGGATCCCTACTTTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
Db GAATTCGCTGATGACAAAGCAGGATCCCTACTTTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2227
2161 TCATCTACAGATTTGGAGAAATGAGTGTGATTTTTCGTACACCCCAAGGTTTGAACCA 2220
Db TCATCTACAGATTTGGAGAAATGAGTGTGATTTTTCGTACACCCCAAGGTTTGAACCA 2287
2281 ACAGTAAATGCTGAGATTTGGAATTAAGGACATTAATTAAGGACATTTTGAAGGACATTTTGAAG 2280
Db ACAGTAAATGCTGAGATTTGGAATTAAGGACATTAATTAAGGACATTTTGAAGGACATTTTGAAG 2347
2341 ACATCAGACATCTGCTTGTAGGATATATAAAGTAAAGGACATTTTGAAGGACATTTTGAAGGAC 2340
Db ACATCAGACATCTGCTTGTAGGATATATAAAGTAAAGGACATTTTGAAGGACATTTTGAAGGAC 2407
2401 TTAGATCAAAACCTGAGTCAACCAATTTTCCCAAAACAGAGTATTTTGAAGGACATTTTGAAGGAC 2400
Db TTAGATCAAAACCTGAGTCAACCAATTTTCCCAAAACAGAGTATTTTGAAGGACATTTTGAAGGAC 2467
2461 ACTACCAAAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2460
Db ACTACCAAAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2527
2521 ATATCTTAAGNAGCTGTTTTAA 2481
Db ACCTCTAAGAGAGCTGTTTTAA 2548

RESULT 3

AAK94164

ID AAK94164 standard; cDNA; 1664 BP.

XX AAK94164;

AC AAK94164;

XX 06-NOV-2001 (first entry)

DT

XX

1801	DB	1801	GTAATGCTTTGGTGGGAATTATGGCTATACAGATAGCTGGGAATGGCTAAATAGAACTACGTTG	1860
1861	QY	1861	GTGGATAATAACACCTCGAAATAACAGCCACATAGCACTGGTGGGAAAAGCTATGCTCTCT	1920
1861	DB	1861	GTGGATAATAACACCTCGAAATAACAGCCACATAGCACTGGTGGGAAAAGCTATGCTCTCT	1920
1921	QY	1921	AATGAAACAGCAGCGCTATAAAAATCATGAGAGACTCTAGATGATAGTATGTTTGGGTTATT	1980
1921	DB	1921	AATGAAACAGCAGCGCTATAAAAATCATGAGAGACTCTAGATGATAGTATGTTTGGGTTATT	1980
2041	QY	2041	TTTGAGGGGTTATGGCTATCTGGTGATGATATCAACAAATTTCTCTGGATGGTTAGG	2040
2041	DB	2041	TTTGAGGGGTTATGGCTATCTGGTGATGATATCAACAAATTTCTCTGGATGGTTAGG	2040
2101	QY	2041	ATAGCTGAAGGAGAAACATCCCAAGAGCATTCGGGAAAGTGACTATTTTACCCCCACAGGGA	2100
2101	DB	2041	ATAGCTGAAGGAGAAACATCCCAAGAGCATTCGGGAAAGTGACTATTTTACCCCCACAGGGA	2100
2101	QY	2101	GAAATTCGGTGATGACCAAGCAGGATCCCTACTTGTGTGAATTCGCTTATGCTATAAAATG	2160
2101	DB	2101	GAAATTCGGTGATGACCAAGCAGGATCCCTACTTGTGTGAATTCGCTTATGCTATAAAATG	2160
2161	QY	2161	TCATACTACAGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGGTTTTCACCGA	2220
2161	DB	2161	TCATACTACAGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGGTTTTCACCGA	2220
2221	QY	2221	ACACGTAATCTGTAGATTGGAATTAAGGACATTAATTCAAACATTTTCGGAAGAGCGCTTT	2280
2221	DB	2221	ACACGTAATCTGTAGATTGGAATTAAGGACATTAATTCAAACATTTTCGGAAGAGCGCTTT	2280
2281	QY	2281	ACATCAGAACACTGGCTTTGTAGGATATATAAAGTAAAGGCCTGTATAACAGGAGACA	2340
2281	DB	2281	ACATCAGAACACTGGCTTTGTAGGATATATAAAGTAAAGGCCTGTATAACAGGAGACA	2340
2341	QY	2341	TTAGATCAAAACCTTCGAGTCACCAACATTTCCCAAAAACAGAAGTATTTGTCAAAAG	2400
2341	DB	2341	TTAGATCAAAACCTTCGAGTCACCAACATTTCCCAAAAACAGAAGTATTTGTCAAAAG	2400
2401	QY	2401	ACTACCAAAAGGACGCTGGCTACATTAATAATTAAGCTGGTTTTTAAGAAAGGCAAGAA	2460
2401	DB	2401	ACTACCAAAAGGACGCTGGCTACATTAATAATTAAGCTGGTTTTTAAGAAAGGCAAGAA	2460
2461	QY	2461	ATATCTAAGAAGACTGTTTAA	2481
2461	DB	2461	ATATCTAAGAAGACTGTTTAA	2481

RESULT 2

ADD94785

ID ADD94785

XX

AC ADD94785;
VV

29-JAN-2004 (first entry)

XX
19 1002-440-22 (1700 2511)

DE Mouse SIMP cDNA sequence.

XX source of immunodominant MHC-associated peptide; STMP; MHC;
KW major histocompatibility complex; human leukocyte antigen, HLA;
KW cytostatic; immunosuppressive; antineoplastic; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW haematological cancer; breast cancer; melanomas; pancreatic cancer;
KW testicular cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; STMP-derived peptide; mouse;
KW murine; gene; ss.

XX

OS Mus musculus.

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

PN WO2003054008-A2.
yy

03-JUL-2003

[illegible]

PF 18-DEC-2002;

20-DEC-2001; 2001US-00028384.
(COMP-) COMPATIGENE INC.
Perreault C, McBride K;
WPI; 2003-559122/52.
P-PSDB; ADD94786.
New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.
Disclosure; SEQ ID NO 3; 66pp; English.
This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC)-associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestine cancer, sarcomas, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the mouse SIMP cDNA which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from GenBank.

Sequence 2710 BP; 678 A; 648 C; 689 G; 695 T; 0 U; 0 Other;

```
Query Match      83.8%; Score 2078.6; DB 10; Length 2710;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 2338; Conservative 0; Mismatches 234; Indels 9; Gaps 1;
```

Qy	1	ATGGCGAGCCCTCGGCCCGGAGAGCAAGTCGTCCTCAACTCGTCCCGCTGG	60
Db	77	ATGGCGAGCCCTCGGCCCGGAGAGCAAGTCGTCCTCAACTCGTCCCGCTGG	136
Qy	61	AGTGGCCTCATGCGCCTGGGAAACAGCGCGACGGCCACACGGCGCCGGGGCCAGTGC	120
Db	137	AGCGGCTCATGCTCTTGGGGAAACAGCCGCCACGGGACATGTGGCGCGGAAACAGAGC	196
Qy	121	GCGCACAAGCGCGGGCGCGCGGCCGCCGAAACCGGCCCGCGCGGGCGTGTCCGGG	180
Db	197	GGGTCCAGGGG-----GCGCGCCGAAAGCGGGGCGCCCGCGGGGTGTCCCGG	247
Qy	181	GGGCTGTGCGAGCGGGCTGGGTGGCAGTGCCTTCTCGCTTCCATCATCTCTTCTCGCC	240
Db	248	GGCTTGTGCGAGCGGGCGGGGTGGCAGTGTCTTCTCTTCCATCATCTCTTCTCGGCC	307
Qy	241	TGGCTTGGCGGCTTCAGCTCGCGGCTTTCGCGGTTCATCCGCTTGGAAAGCATCATCCAC	300
Db	308	TGGCTGGCGGCTTCAGCTCGCGGCTTTCGCGGTTCATCCGCTTCGAGAGCATCATCCAC	367
Qy	301	GAGTTTCGACCGGTGTTAACTATAGATCAACACATCATCTTGCATCTCATTGGGTCTAT	360
Db	368	GAGTTTCGACCGGTGTTAACTATAGATCAACACATCATCTTGCATCTCATTGGATTTCTAT	427
Qy	361	GAATTTTAAATTTGTTTGAATGAAGAGCATGGTATCCCACTAGGAAGATATAGTAGTGGT	420
Db	428	GAGTTTCAAAATTTGTTTGAATGAAGAGCATGGTATCCCACTCGGAAGATATAGTGGGTGGC	487
Qy	421	ACTGTTTACCGAGGTTGATGAATAACCGCTTGGCCTTATTCATTGGATTTTAAATACATTG	480
Db	488	ACCGTTTACCGAGGTTGATGAATGAAGAGCATGGGCTTATTCATTGGATTTTAAATACATTG	547

termed source of immunodominant major histocompatibility complex (MHC) - associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestine cancer, sarcomas, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the human SIMP cDNA which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from GenBank.

SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Query Match 100.0%; Score 2481; DB 10; Length 2481;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGAGCGCTCGGCCCGGAGACAGCAAGTCTCCCTCAACTCGTCCCGCTGG 60
 Db 1 ATGCGAGCGCTCGGCCCGGAGACAGCAAGTCTCCCTCAACTCGTCCCGCTGG 60

QY 61 AGTGGCCTCATGGCCCTGGGAAACAGCGGACAGCCAGCGGCCCGGGCCCAAGTGC 120
 Db 61 AGTGGCCTCATGGCCCTGGGAAACAGCGGACAGCCAGCGGCCCGGGCCCAAGTGC 120

QY 121 GCGCACAGGCGGCG 180
 Db 121 GCGCACAGGCGGCG 180

QY 181 GGGCTGTGCGAGCGCGGTGGGTGGCGAGTGGCTTCTCTCTCCCTCACCATCTCTTCTCGGCC 240
 Db 181 GGGCTGTGCGAGCGCGGTGGGTGGCGAGTGGCTTCTCTCTCCCTCACCATCTCTTCTCGGCC 240

QY 241 TGGCTTCCCGCTCAGCTCGCGCTCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCT 300
 Db 241 TGGCTTCCCGCTCAGCTCGCGCTCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCT 300

QY 301 GAGTTCGACCGCGTGTAACTATAGATCAACATCATCTTGCATCTCATGGGTCTAT 360
 Db 301 GAGTTCGACCGCGTGTAACTATAGATCAACATCATCTTGCATCTCATGGGTCTAT 360

QY 361 GAATTTTAAATTTGGTTTGAAGAGCATGTATCCACTAGGAAGATAGTGGT 420
 Db 361 GAATTTTAAATTTGGTTTGAAGAGCATGTATCCACTAGGAAGATAGTGGT 420

QY 421 ACTGTTTACCAGGTTGATGATAACCGCTGGCTTATTTCATTGGATTTTAAATACATTG 480
 Db 421 ACTGTTTACCAGGTTGATGATAACCGCTGGCTTATTTCATTGGATTTTAAATACATTG 480

QY 481 AACATCTCTTACATAGAGCATGTATGTGTCTTGTGCTTGTGCTTGTGCTTGTGCTT 540
 Db 481 AACATCTCTTACATAGAGCATGTATGTGTCTTGTGCTTGTGCTTGTGCTTGTGCTT 540

QY 541 ACATCTATATCTACTTTTCCCTGCTTACAAGAGAACTTTTGGAAACCAAGGACAGGACTTTTA 600
 Db 541 ACATCTATATCTACTTTTCCCTGCTTACAAGAGAACTTTTGGAAACCAAGGACAGGACTTTTA 600

QY 601 GCTGCTTGTATTTATGCTATGTACAGGCTACATATCTCGCTCAGTACGTGGTCTT 660
 Db 601 GCTGCTTGTATTTATGCTATGTACAGGCTACATATCTCGCTCAGTACGTGGTCTT 660

QY 661 GATAATGAAGGCATTTGCTATTTTGGCACTTCAGTTCACATCTATTTATGGTAAATCT 720
 Db 661 GATAATGAAGGCATTTGCTATTTTGGCACTTCAGTTCACATCTATTTATGGTAAATCT 720

QY 721 GTAAATACTGGTTCAGTTTTTTGGACAAATGTGCTGCTTATCTTATATGTC 780

Db 721 GTAAATACTGGTTCAGTTTTTTGGACAAATGTGCTGCTTATCTTATATGTC 780
 QY 781 TCTGCTTGGGGTGGTTATGTTATTTATCATCAATCTTATTCACATGCATGTTGTTG 840
 Db 781 TCTGCTTGGGGTGGTTATGTTATTTATCATCAATCTTATTCACATGCATGTTGTTG 840

QY 841 TTACTGATCAGAGATACAGCAAGAGCTTACATAGCATATAGCATTCTTACATGTCG 900
 Db 841 TTACTGATCAGAGATACAGCAAGAGCTTACATAGCATATAGCATTCTTACATGTCG 900

QY 901 GGTTTAATATTATCAATGCAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTCAA 960
 Db 901 GGTTTAATATTATCAATGCAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTCAA 960

QY 961 CACATCGCAGCTGCGAGTCTTTCATTTGCTGCAAGCTTATGCTTCTTCAGTATCTG 1020
 Db 961 CACATCGCAGCTGCGAGTCTTTCATTTGCTGCAAGCTTATGCTTCTTCAGTATCTG 1020

QY 1021 AGAGACCGATTAAACAAAACAGAGTTCAGACCCCTTTTCTTTTGGGTATCACTAGCT 1080
 Db 1021 AGAGACCGATTAAACAAAACAGAGTTCAGACCCCTTTTCTTTTGGGTATCACTAGCT 1080

QY 1081 GCAGGTGCTGTGTCCTTAGTCTCATCTATTGACTTATACAGGTTACATTTGCACCATG 1140
 Db 1081 GCAGGTGCTGTGTCCTTAGTCTCATCTATTGACTTATACAGGTTACATTTGCACCATG 1140

QY 1141 AGTGGCAGGTTTATTTCAATTTGGGATACGGGTATGCAAAAATACACATTTCCAATTA 1200
 Db 1141 AGTGGCAGGTTTATTTCAATTTGGGATACGGGTATGCAAAAATACACATTTCCAATTA 1200

QY 1201 GCATCAGTCTGAGCATCAACTAGCTTGGGTGCTTCTTCTTTGATCTACATATT 1260
 Db 1201 GCATCAGTCTGAGCATCAACTAGCTTGGGTGCTTCTTCTTTGATCTACATATT 1260

QY 1261 CTTGTATGACCTTCCAGCAGCCCTTTGGTTCGCATCAAAAATATCAACGATGAAAGA 1320
 Db 1261 CTTGTATGACCTTCCAGCAGCCCTTTGGTTCGCATCAAAAATATCAACGATGAAAGA 1320

QY 1321 GTATTTTGTGCTCTATATGCAATCAGTCTGCTACTTTGCTGGAGTATGCTGGAGCTG 1380
 Db 1321 GTATTTTGTGCTCTATATGCAATCAGTCTGCTACTTTGCTGGAGTATGCTGGAGCTG 1380

QY 1381 ATGTTGACCTTGACTCCAGTCTGTGTATGTGCTGCAATTTGCTTTTCAAAATGTTTT 1440
 Db 1381 ATGTTGACCTTGACTCCAGTCTGTGTATGTGCTGCAATTTGCTTTTCAAAATGTTTT 1440

QY 1441 GAGCCTATTTGGGGATGACATGAAGGGAATCCACCTGTGGAGGACAGCAGTGTAT 1500
 Db 1441 GAGCCTATTTGGGGATGACATGAAGGGAATCCACCTGTGGAGGACAGCAGTGTAT 1500

QY 1501 GAGGATGCAAAAGAAACCAAGGAAATTTGTATGATAGGCGAGGTAAAGTGAAGAAAT 1560
 Db 1501 GAGGATGCAAAAGAAACCAAGGAAATTTGTATGATAGGCGAGGTAAAGTGAAGAAAT 1560

QY 1561 GCACTGCAACAGGAAATCTCAAGGAGTTAGGCCCTTAATATAAAAGCATTTGTCACC 1620
 Db 1561 GCACTGCAACAGGAAATCTCAAGGAGTTAGGCCCTTAATATAAAAGCATTTGTCACC 1620

QY 1621 ATGTTGATCTGATGCTATTGATGATTTGCTGCCACTGTACTGGGTCAACAGCAAT 1680
 Db 1621 ATGTTGATCTGATGCTATTGATGATTTGCTGCCACTGTACTGGGTCAACAGCAAT 1680

QY 1681 GCCTACTCTAGTCCAGTGTAGTCTTGCCTCATACATCATGATGGCACCAGGAATATC 1740
 Db 1681 GCCTACTCTAGTCCAGTGTAGTCTTGCCTCATACATCATGATGGCACCAGGAATATC 1740

QY 1741 TTAGATGATTTTAGAGAGCTTATCTTTTGGCTTAAGGCAAAAATACAGATGAACATGCA 1800
 Db 1741 TTAGATGATTTTAGAGAGCTTATCTTTTGGCTTAAGGCAAAAATACAGATGAACATGCA 1800

QY 1801 GTAAATCTTGGTGGATTTATGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG 1860

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 20:50:43 ; Search time 1138.18 Seconds

(Without alignments)
11442.663 Million cell updates/sec

Title: US-10-028-384-1

Perfect score: 2481

Sequence: 1 atggcggagcctcggccccc.....tatctaagaagactgttttaa 2481

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8259772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Geneseqn23Sep04:*
- 2: Geneseqn1380s:*
- 3: Geneseqn1390s:*
- 4: Geneseqn2000s:*
- 5: Geneseqn2001as:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2481	100.0	2481	10	ADD94783 Human SIM
2	2078.6	83.8	2710	10	ADD94785 Mouse SIM
3	1328.8	53.6	1664	4	AAK94164 Human ful
4	1328.8	53.6	1664	12	ADL30661 Full leng
5	1061	42.8	1203	4	AAK33284 Human col
6	1061	42.8	1209	6	ABL98850 Human pol
7	1060.4	42.7	1543	2	AAK85055 Human sec
8	1060.4	42.7	1543	8	ACD18981 Novel hum
9	1060.4	42.7	1543	12	ADG78372 Human sec
10	1060.4	42.7	1543	12	ADN60663 Human sec
11	972	39.2	2546	2	AAV44866 Clone CT5
12	972	39.2	2546	5	AAK98453 Human CDN
13	955.8	38.5	2417	10	ADD94789 Drosophil
14	955.8	38.5	2699	4	ABL13247 Drosophil
15	954	38.0	2510	5	AAK93772 Human CDN
16	892.4	36.0	2537	4	AAK08289 Human sec
17	892	36.0	2547	4	AAK08315 Human sec
18	892	36.0	2660	6	ABQ54750 Human ova
19	867.2	35.0	1114	4	AAH99794 Human pro
20	861.4	34.7	2284	4	AAH18021 Human CDN
21	781.6	30.7	4922	4	ABL13246 Drosophil

ALIGNMENTS

RESULT 1

ADD94783
ID ADD94783 standard; CDN; 2481 BP.

XX AC ADD94783;

DT 23-JAN-2004 (first entry)

XX DE Human SIMP CDNA sequence.

XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;
XX KW major histocompatibility complex; human leukocyte antigen; HLA;
XX KW cytotoxic; immunosuppressive; antisense therapy; gene therapy; cancer;
XX KW lung cancer; intestine cancer; sarcoma; prostate cancer;
XX KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
XX KW haematological cancer; immune response; lymphoid cell proliferation;
XX KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
XX KW gene; ss.
XX OS Homo sapiens.
XX PN WO2003054008-A2.
XX PD 03-JUL-2003.
XX PF 18-DEC-2002; 2002WO-CA001967.
XX PR 20-DEC-2001; 2001US-00028384.
XX PA (COMP-) COMPATIGENE INC.
XX PI Perreault C, McBride K;
XX DR WPI: 2003-559122/52.
XX DR P-PSDB; ADD94784.

New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.

Claim 6; SEQ ID NO 1; 66pp; English.

This invention relates to a novel isolated or purified human protein.

626218240Fl NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4732256 5', mRNA sequence.
 BG620025
 BG620025.1 GI:136711396
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 776)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L10CM1591 row: 6 column: 09
 High quality sequence spot: 727.

FEATURES
source

RESULT 15	ACCESSION	REFERENCE
CD103829	VERSION	AUTHORS
LOCUS	KEYWORDS	TITLE
DEFINITION	SOURCE	JOURNAL
	ORGANISM	COMMENT

ORIGIN

Query Match	20.8%;	Score 517;	DB 4;	Length 776;
Best Local Similarity	99.7%;	Pred. No. 1e-266;		
Matches 617;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
1663	GGATAATAACACCTGGNAATAACAGCCACATAGCACTGGTGGGAAAGCTATGCTCTCAA	1922		
Ddb	1	GGATAATAACACCTGGAAATAACAGCCACATAGCTGGTGGGAAAGCTATGCTCTCCAA	60	
1923	TGAAACAGCAGCCCTATATAAATCATGAGGACTCTAGATCTAGATTATGTTTGGTTATTTT	1982		
Ddb	61	TGAAACAGCAGCCCTATATAAATCATGAGGACTCTAGATCTAGATTATGTTTGGTTATTTT	120	
1993	TGGAGGGTTATTGGCTATTCTGGTGATGATATCAACAAATTTCTCGGATGGTTAGGAT	2042		
Ddb	121	TGGAGGGGTTATTGGCTATTCTGGTGATGATATCAACAAATTTCTCGGATGGTTAGGAT	180	
2043	AGCTGAAGGAGNACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCCAACAGGGAGA	2102		
Ddb	181	AGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCCAACAGGGAGA	240	
2103	ATTCGGTGTTAGACAAAGCAGGATCCCTACTTTGTTGAAATTCGCTTATGATATAAATGTC	2162		
Ddb	241	ATTCGGTGTTAGACAAAGCAGGATCCCTACTTTGTTGAAATTCGCTTATGATATAAATGTC	300	
2163	ATACTACAGATTTGGAGAAATCAGCTGGATTTTCGTACACCCCAACAGGTTTTGACCGAAC	2222		
Db	301	ATACTACAGATTTGGAGAAATCAGCTGGATTTTCGTACACCCCAACAGGTTTTGACCGAAC	360	

ORIGIN

Query Match

20.5%; Score 508; DB 6; Length 793;

QY 1794 TGCACGAGTAAATGCTTTGGGATATGCTATCAGATAGCTGGAAATGGCTAATAGAAC 1853
 Db 361 TGCACGAGTAAATGCTTTGGGATATGCTATCAGATAGCTGGAAATGGCTAATAGAAC 420
 QY 1854 TACCTTGGTGAATATACACCTGGATACACCCACATAGACACTGGTGGGAAAGCTAT 1913
 Db 421 TACCTTGGTGAATATACACCTGGATACACCCACATAGACACTGGTGGGAAAGCTAT 480
 QY 1914 CTCCTTCAATGAACACACGACCTATAAAATCATGAGAGCTCTAGATGTAGATATGTTT 1973
 Db 481 CTCCTTCAATGAACACGACCTATAAAATCATGAGAGCTCTAGATGTAGATATGTTT 540
 QY 1974 GGTATATTT 1982
 Db 541 GGTATATTT 549
 RESULT 12
 BG434322 627 bp mRNA linear EST 14-MAR-2001
 LOCUS 602506307FL NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4603584 5',
 DEFINITION mRNA sequence.
 ACCESSION BG434322.1 GI:13340828
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 627)
 NIH-MGC <http://mgi.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LICM1337 row: n column: 01
 High quality sequence stop: 625.
 Location/Qualifiers
 1..627
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clones="IMAGE:4603584"
 /lab_hosts="DH10B (T1 phage-resistant)"
 /clone_lib="NIH MGC 79"
 /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
 Site 1: SfiI (ggcgctcgcc); Site 2: SfiI
 as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.3
 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."
 ORIGIN
 Query Match 22.0%; Score 546; DB 4; Length 627;
 Best Local Similarity 100.0%; Pred. No. 2.2e-282;
 Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 621 TGTACCGGCTACATATCTGGTCAGTAGCTGGATCCTTTGATAATGAAGGATGCTAT 680
 Db 81 TGTACCGGCTACATATCTGGTCAGTAGCTGGATCCTTTGATAATGAAGGATGCTAT 140

QY 681 TTTTGCACTTCAGTTCCATACACTATTTATGGGTAATAATCTGTAATAAACTGGGTTCAGTTT 740
 Db 141 TTTTGCACTTCAGTTCCATACACTATTTATGGGTAATAATCTGTAATAAACTGGGTTCAGTTT 200
 QY 741 TTGGCAATGTCCTGCTGCTTATCCCTATTCCTATTTCTATATGCTCTCTCTGGGTGGTTATGT 800
 Db 201 TTGGCAATGTCCTGCTGCTTATCCCTATTCCTATTTCTATATGCTCTCTCTGGGTGGTTATGT 260
 QY 801 ATTTATCATCAATCTTTATTCCTACTCCATGTATTTCTGTTTACTGTATGCAGATACAG 860
 Db 261 ATTTATCATCAATCTTTATTCCTACTCCATGTATTTCTGTTTACTGTATGCAGATACAG 320
 QY 861 CAAAAGAGTCTACATAGCATATAGACATTTCTACATTTGGGTTTAAATATTAATCAATGCA 920
 Db 321 CAAAAGAGTCTACATAGCATATAGACATTTCTACATTTGGGTTTAAATATTAATCAATGCA 380
 QY 921 GATACCTTTTGTGGGATCCAGCCCAATCAGAACAGTGAACACATGCGAGTGCAGGTGT 980
 Db 381 GATACCTTTTGTGGGATCCAGCCCAATCAGAACAGTGAACACATGCGAGTGCAGGTGT 440
 QY 981 CTTTGCACTTCAGTCAAGCTTATGCTTTCTTGAGTATCTGAGAGACCGATTAAACAAACA 1040
 Db 441 CTTTGCACTTCAGTCAAGCTTATGCTTTCTTGAGTATCTGAGAGACCGATTAAACAAACA 500
 QY 1041 AGAGTTCCAGACCCCTTTCTTTTGGGTATFCACTAGCTGCAGTGTCTGTTCCTTAG 1100
 Db 501 AGAGTTCCAGACCCCTTTCTTTTGGGTATFCACTAGCTGCAGTGTCTGTTCCTTAG 560
 QY 1101 TGTCACTATTTGACTTATACAGGTTTACATTCACCATGGAGTGCAGGTTTATTCATT 1160
 Db 561 TGTCACTATTTGACTTATACAGGTTTACATTCACCATGGAGTGCAGGTTTATTCATT 620
 QY 1161 GTGGGA 1166
 Db 621 GTGGGA 626
 RESULT 12
 CN263069 577 bp mRNA linear EST 16-MAY-2004
 LOCUS 17000600180485 GRN_PRENU Homo sapiens cDNA 5', mRNA sequence.
 DEFINITION CN263069
 ACCESSION CN263069.1 GI:47279483
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 577)
 Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J.,
 Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,
 Lebkowski, J. and Stanton, L.W.
 Transcription characterization elucidates signaling networks that
 control human ES cell growth and differentiation
 Nat. Biotechnol. 22 (6), 707-716 (2004)
 Contact: Brandenberger R
 Regenerative Medicine
 Gen Corporation
 230 Constitution Drive, Menlo Park, CA 94025, USA
 Tel: 650 473 8658
 Fax: 650 473 7760
 Email: rbrandenberger@gen.com
 Insert length: 577, Std Error: 0.00.
 Location/Qualifiers
 1..577
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic stem cell, retinoic acid and
 mitogen-treated hES cell line H7"
 /clone_lib="GRN_PRENU"
 /note="oligo dt primed, full-length enriched cDNA library"

Plate: 5 row: E column: 12
High quality sequence stop: 562.

FEATURES

Source
1. 562
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L11SNU354-5-E12"
/sex="M"
/tissue_type="Liver"
/cell_type="Polyclonal"
/lab_hosts="Topi0r"
/clone_lib="L11SNU354"
/notes="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI; Site 2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Topi0r by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

ORIGIN

Query Match 22.7%; Score 562; DB 6; Length 562;
Best Local Similarity 100.0%; Pred. No. 5.1e-291;
Matches 562; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1643 TGATGTTGCTGCTCACTGACCTGGGTGTCACAGCAATGCTTCTAGTCCAAAGTGTAG 1702
DB 1 TGATGTTGCTGCTCACTGACCTGGGTGTCACAGCAATGCTTCTAGTCCAAAGTGTAG 60

QY 1703 TCCTGGGCTCATACAATCATGATGGCCACAGGAATATCTTAGATGATTTTAGAGAAGCTT 1762
DB 61 TCCTGGGCTCATACAATCATGATGGCCACAGGAATATCTTAGATGATTTTAGAGAAGCTT 120

QY 1763 ACTTTGGCTAAGCAAAATACAGATGAACATGCCAGATATGCTTGGTGGGATTATG 1822
DB 121 ACTTTGGCTAAGCAAAATACAGATGAACATGCCAGATATGCTTGGTGGGATTATG 180

QY 1823 GCTATCAGATAGCTGGGAATGCTTAATAGAACTACGTTGGTGATATACACCTGGAATA 1882
DB 181 GCTATCAGATAGCTGGGAATGCTTAATAGAACTACGTTGGTGATATACACCTGGAATA 240

QY 1883 ACAGCCATAGCACTGGTGGGAAAAGCTATGCTTCTTAATGAACAGAGCCTATAAAA 1942
DB 241 ACAGCCATAGCACTGGTGGGAAAAGCTATGCTTCTTAATGAACAGAGCCTATAAAA 300

QY 1943 TCATGAGGACTCTAGATGTAGATATGCTTTGGTATATTTTGGAGGGTTATTGGCTATT 2002
DB 301 TCATGAGGACTCTAGATGTAGATATGCTTTGGTATATTTTGGAGGGTTATTGGCTATT 360

QY 2003 CTGCTGATGATATCAACAAATTTCTCTGGATGTTAGGATAGCTGAAGGAGAACTCCCA 2062
DB 361 CTGCTGATGATATCAACAAATTTCTCTGGATGTTAGGATAGCTGAAGGAGAACTCCCA 420

QY 2063 AAGACATTTGGGAAAAGTGACTATTTTACCCACAGGGAGAAATTCCTGTGTAGCAAGGAG 2122
DB 421 AAGACATTTGGGAAAAGTGACTATTTTACCCACAGGGAGAAATTCCTGTGTAGCAAGGAG 480

QY 2123 GATCCCTACTTTTGGATTTGCTTATGTATPAAAAATGCTATACAGATTTGGAGAAA 2182
DB 481 GATCCCTACTTTTGGATTTGCTTATGTATPAAAAATGCTATACAGATTTGGAGAAA 540

QY 2183 TGCAGCTGGATTTTTCGTACACC 2204

Db 541 TGCAGCTGGATTTTTCGTACACC 562

RESULT 10

BG698273 735 bp mRNA linear EST 07-MAY-2001
602650188F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4803332 5',
LOCUS mRNA sequence.

DEFINITION

ACCESSION BG698273.1 GI:13965393

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: James Cleaver, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA

Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LMNL at:

http://image.lmnl.gov

Plate: LLAM10698 row: p column: 21

High quality sequence start: 2

High quality sequence stop: 734.

Location/Qualifiers

1. 735

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4803332"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NCI CGAP Skn3"

/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. primer: Oligo dt.

Average insert size 1.5kb. Library constructed by Life

Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 22.1%; Score 549; DB 4; Length 735;
Best Local Similarity 100.0%; Pred. No. 5.4e-284;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1434 TGTTTTGAGCACTATTTGGGGATGACATGAAGGGAATCCACCTGTGGAGGACAG 1493
DB 1 TGTTTTGAGCACTATTTGGGGATGACATGAAGGGAATCCACCTGTGGAGGACAG 60

QY 1494 CAGTGTAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGCAGGTAAGTGAG 1553
DB 61 CAGTGTAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGCAGGTAAGTGAG 120

QY 1554 GACATGCAACTGACAGGAAACCTGAAGGGATTAGCCCTATATAAAGCAT 1613
DB 121 GAAACATGCAACTGACAGGAAACCTGAAGGGATTAGCCCTATATAAAGCAT 180

QY 1614 TGTCAACATGTTGCTGCTGATGCTATTGATGATGTTTCTGTCACCTGTACCTGGTGCAC 1673
DB 181 TGTCAACATGTTGCTGCTGATGCTATTGATGATGTTTCTGTCACCTGTACCTGGTGCAC 240

QY 1674 AGCAATGCTACTCTAGTCCAAAGTGTAGTCTGSCCTCATACATCATGATGCACACAG 1733
DB 241 AGCAATGCTACTCTAGTCCAAAGTGTAGTCTGSCCTCATACATCATGATGCACACAG 300

QY 1734 GAATATCTTAGATGATTTTAGAAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACA 1793
DB 301 GAATATCTTAGATGATTTTAGAAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACA 360

QY 1678 AATGCTTACTTATGTCACAGTGTAGTCTGGGCTCATACATCATGATGCGACAGGAAT 1737
 Db 241 AATGCTTACTTATGTCACAGTGTAGTCTGGGCTCATACATCATGATGCGACAGGAAT 300
 QY 1738 ATCTTAGATGATTTAGAGAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAATGCA 1797
 Db 301 ATCTTAGATGATTTAGAGAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAATGCA 360
 QY 1798 CGAGTAATGCTTGGTGGGATTTAGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGC 1857
 Db 361 CGAGTAATGCTTGGTGGGATTTAGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGC 420
 QY 1858 TTGGTGATTAATACACCTGGATTAACAGCAATACAGCAATAGCACTGGTGGGAAAGCTATGCT 1917
 Db 421 TTGGTGATTAATACACCTGGATTAACAGCAATACAGCAATAGCACTGGTGGGAAAGCTATGCT 480
 QY 1918 TCTAATGAACAGCAGCCTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTT 1777
 Db 481 TCTAATGAACAGCAGCCTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTT 540
 QY 1978 ATTTTCGAGGGTATTGGCTATTCTGGTGATGATATCAAC 2019
 Db 541 ATTTTCGAGGGTATTGGCTATTCTGGTGATGATATCAAC 582

RESULT 8
 BQ877881 889 bp mRNA linear EST 16-AUG-2002
 LOCUS AGENCOURT 7981106 LuPSki dorsal_root_ganglion Homo sapiens cDNA
 DEFINITION clone IMAGE:6185364 5', mRNA sequence.

ACCESSION BQ877881
 VERSION BQ877881.1 GI:22269889
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 889)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rcapbs@mail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM13576 row: a column: 13
 High quality sequence stop: 673.

FEATURES
 source
 Location/Qualifiers
 1..889
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6185364"
 /sex="male"
 /tissue_type="dorsal root ganglia"
 /dev_stage="adult, 36 yr"
 /lab_host="DH10B"
 /clone_lib="LuPSki_dorsal_root_ganglion"
 /note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
 Note1: Site 2: SalI; cDNA made by oligo-dt priming.
 Directionally cloned using the following adaptors:
 5'-TCGACCCACGCGCCG-3' and
 5'-GATAGTCTAGTCCGCGCGCCGCT(15)-3'. Size selected >
 1 kb for average insert length 1.7 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine) and is available through Life

Technologies."

ORIGIN
 Query Match 23.5%; Score 582; DB 5; Length 889;
 Best Local Similarity 100.0%; Pred. No. 8.5e-302;
 Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1900 GTGGGAAAAGCTATGCTTTCTTAATGAAAACAGCAGCCTATAAAATCATGAGGACTCTAGAT 1959
 Db 1 GTGGGAAAAGCTATGCTTTCTTAATGAAAACAGCAGCCTATAAAATCATGAGGACTCTAGAT 60
 QY 1960 GTAGATTATGTTTGGTTTATTTTGGAGGGGTATTGGCTATTCTGGTGATGATATCAAC 2019
 Db 61 GTAGATTATGTTTGGTTTATTTTGGAGGGGTATTGGCTATTCTGGTGATGATATCAAC 120
 QY 2020 AATTTCTCTGATGGTTAGGATAGCTGAAGAGAAACATCCCAAGACATTCGGGAAAGT 2079
 Db 121 AATTTCTCTGATGGTTAGGATAGCTGAAGAGAAACATCCCAAGACATTCGGGAAAGT 180
 QY 2080 GACTATTTTACCCCAAGGAGAAATTCGGTGTAGACAAAGCAGGATCCCTACTTTGTTG 2139
 Db 181 GACTATTTTACCCCAAGGAGAAATTCGGTGTAGACAAAGCAGGATCCCTACTTTGTTG 240
 QY 2140 AATTGCTTATGTATTAATGTCATCTACAGATTGGAGAAATGCGAGCTGGATTTTCGT 2199
 Db 241 AATTGCTTATGTATTAATGTCATCTACAGATTGGAGAAATGCGAGCTGGATTTTCGT 300
 QY 2200 ACACCCCAAGGTTTGGACGAAACAGCTAATGCTGAGATTGGAAATTAAGGACATTAAATTC 2259
 Db 301 ACACCCCAAGGTTTGGACGAAACAGCTAATGCTGAGATTGGAAATTAAGGACATTAAATTC 360
 QY 2260 AAACATTTGGAAGAGCCCTTTACATCAGAACTGCTGGCTTTGTAGGATATATAAGATAAAA 2319
 Db 361 AAACATTTGGAAGAGCCCTTTACATCAGAACTGCTGGCTTTGTAGGATATATAAGATAAAA 420
 QY 2320 GCACCTGATAACAGGAGAGACATTAGATCACAAACCTCGAGTCACCAACATTTTCCAAAA 2379
 Db 421 GCACCTGATAACAGGAGAGACATTAGATCACAAACCTCGAGTCACCAACATTTTCCAAAA 480
 QY 2380 CAGAGATATTTGTCAAAGAGACTACCAAGAGAGCGTGGCTACATTAAAAATTAAGCTG 2439
 Db 481 CAGAGATATTTGTCAAAGAGACTACCAAGAGAGCGTGGCTACATTAAAAATTAAGCTG 540
 QY 2440 GTTTTAAAGAAAGGCAAGAAATATCTAAGAGAGCTGTTTAA 2481
 Db 541 GTTTTAAAGAAAGGCAAGAAATATCTAAGAGAGCTGTTTAA 582

RESULT 9
 BQ128931
 LOCUS K-EST0178489 L11SNU354 Homo sapiens cDNA clone L11SNU354-5-E12 5',
 DEFINITION mRNA sequence.
 ACCESSION BQ128931
 VERSION BQ128931.1 GI:28092254
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 562)
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
 Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
 Kim, Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr

```

RESULT 6
CD358477          903 bp      mRNA      linear      EST 29-MAY-2003
LOCUS              AGENCOURT 14252099 NIH_MGC_180 Homo sapiens cDNA clone
DEFINITION         IMAGE:30388860 5', mRNA sequence.
ACCESSION          CD358477
VERSION            CD358477.1 GI:31129912
KEYWORDS           EST.
SOURCE             Homo sapiens (human)
ORGANISM           Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1 (bases 1 to 903)
AUTHORS           NIH-MGC http://mgc.nci.nih.gov/.
TITLE             National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL            Unpublished (1999)
COMMENT           Contact: Daniela S. Gerhard, Ph.D.
                   Office of Cancer Genomics
                   National Cancer Institute / NIH
                   Bldg. 31 Rm10A07 Bethesda, MD 20892
                   Email: cgaops-remail.nih.gov
                   Tissue Procurement: Dr. Michael Brownstein
                   cDNA Library Preparation: Invitrogen Corp
                   CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                   DNA Sequencing by: Agencourt Bioscience Corporation
                   Clone distribution: MGC clone distribution information can be
                   found through the I.M.A.G.E. Consortium/LLNL at:
                   http://image.llnl.gov
                   Plate: NDAM461 row: g column: 13
                   High quality sequence stop: 641.
FEATURES           Location/Qualifiers
                   1..903
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:30388860"
                   /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                   /clone_lib="NIH_MGC_180"
                   /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
                   Site 2: EcoRV (destroyed); Library is oligo-dT primed and
                   directionally cloned (EcoRV site is destroyed upon
                   cloning). Average insert size 1.68 kb. Library was
                   constructed by (invitrogen). Note: this is a NIH_MGC
                   Library."
ORIGIN
Query Match      23.8%; Score 590; DB 6; Length 903;
Best Local Similarity 100.0%; Pred. No. 4.1e-306;
Matches 590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1872 CACCTGAATAACAGCCACATAGACCTGGTGGGAAAGCTATGCTTCTTATGAACACAGC 1931
DB 60 CACCTGAATAACAGCCACATAGACCTGGTGGGAAAGCTATGCTTCTTATGAACACAGC 119
QY 1932 AGCCTATAAATCATGAGGACTCTAGATGTAGATTATGTTTGGTTATTTTGGAGGGGT 1991
DB 120 AGCCTATAAATCATGAGGACTCTAGATGTAGATTATGTTTGGTTATTTTGGAGGGGT 179
QY 1992 TATTGGCTATCTGGTATGATATCAACAAATTTCTGTGATGGTTAGATAGCTGAAGG 2051
DB 180 TATTGGCTATCTGGTATGATATCAACAAATTTCTGTGATGGTTAGATAGCTGAAGG 239
QY 2052 AGAACATCCCAAGACATTCGGGAAAGTGACATTTTACCCACAGGAGGAAATCCCGTGT 2111
DB 240 AGAACATCCCAAGACATTCGGGAAAGTGACATTTTACCCACAGGAGGAAATCCCGTGT 299
QY 2112 AGACAAAGCAGGATCCCTACTTTTGTGAATTCGCTTATGATATATAAAATGCTACTACAG 2171
DB 300 AGACAAAGCAGGATCCCTACTTTTGTGAATTCGCTTATGATATATAAAATGCTACTACAG 359
QY 2172 ATTTGGAGAAATGCAGCTGGATTTTCGTACACACCCCGAGGTTTTCGACCAACACGTAATGC 2231

```

```

DB 360 ATTTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGGTTTTCACCGAACACGTAATGC 419
QY 2232 TGAGATTGGAAATAGGACATTAAATTCAAACATTTCGAAGAAGCCTTTACATCAGAACA 2291
DB 420 TGAGATTGGAAATAGGACATTAAATTCAAACATTTCGAAGAAGCCTTTACATCAGAACA 479
QY 2292 CTGGCTTGTAGGATATATAAAGTAAAGCACTGATTAACAGGAGACATTTAGATCAAA 2351
DB 480 CTGGCTTGTAGGATATATAAAGTAAAGCACTGATTAACAGGAGACATTTAGATCAAA 539
QY 2352 ACCTCGAGTCACCAACATTTCCCAACACAGAGATTTCGCAAGAGAGACTACCAAAAG 2411
DB 540 ACCTCGAGTCACCAACATTTCCCAACACAGAGATTTCGCAAGAGAGACTACCAAAAG 599
QY 2412 GAAGCGTGCTGCTACATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAAA 2461
DB 600 GAAGCGTGCTGCTACATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAAA 649

RESULT 7
BP871615          582 bp      mRNA      linear      EST 27-JUL-2004
LOCUS              BP871615 Sugano cDNA library, embryonal kidney Homo sapiens cDNA
DEFINITION         clone HKR00887, mRNA sequence.
ACCESSION          BP871615
VERSION            BP871615.1 GI:50716064
KEYWORDS           EST.
SOURCE             Homo sapiens (human)
ORGANISM           Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1 (bases 1 to 582)
AUTHORS           Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J.,
                   Naki, K. and Sugano S.
TITLE             Database of Transcriptional Start Sites (DBTSS) for Comparative
                   Studies of the Promoters of Human and Mouse Genes
JOURNAL            Unpublished (2004)
COMMENT           Contact: Yutaka Suzuki
                   Department of Virology
                   Institute of Medical Science, University of Tokyo
                   4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
                   Email: yusuzuki@ims.u-tokyo.ac.jp.
FEATURES           Location/Qualifiers
                   1..582
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="HKR00887"
                   /tissue_type="embryonal kidney"
                   /cell_line="293"
                   /clone_lib="Sugano cDNA library, embryonal kidney"
ORIGIN
Query Match      23.5%; Score 582; DB 5; Length 582;
Best Local Similarity 100.0%; Pred. No. 8.1e-302;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1438 TTTTGACACTATTTGGGGATGACATGAAGAAGGAAATCCACCTGTGGAGGACAGCAGT 1497
DB 1 TTTTGACACTATTTGGGGATGACATGAAGAAGGAAATCCACCTGTGGAGGACAGCAGT 60
QY 1498 GATGAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGAGGAGTAAAGTGAAGAA 1557
DB 61 GATGAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGAGGAGTAAAGTGAAGAA 120
QY 1558 CATGCACTGACAGCAAGAAACACTGAGAGGATTTAGCCCTATATAAAGCATGTC 1617
DB 121 CATGCACTGACAGCAAGAAACACTGAGAGGATTTAGCCCTATATAAAGCATGTC 180
QY 1618 ACCATGTTGATGCTGATGCTATTGATGATGTTTGGTGTCCACTGTACCTGGGTCAAGC 1677
DB 181 ACCATGTTGATGCTGATGCTATTGATGATGTTTGGTGTCCACTGTACCTGGGTCAAGC 240

```



```

Db      507 TCATATACAGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCACCGA 566
QY      2221 ACAGTAATCTCAGATGGAAATAGGACATTAATTCARACATTTGGAGAGCCCTT 2280
Db      567 ACAGTAATCTCAGATGGAAATAGGACATTAATTCARACATTTGGAGAGCCCTT 626
QY      2281 ACATCAGAACACTGGCTT 2298
Db      627 ACATCAGAACACTGGCTT 644

RESULT 3
LOCUS   BX335717 1057 bp mRNA linear EST 08-APR-2004
DEFINITION BX335717 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0D1018YD04 5-PRIME, mRNA sequence.
ACCESSION BX335717
VERSION   BX335717.2 GI:46281069
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS   Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On May 2, 2003 this sequence version replaced gi:30345448.
          Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 EVRY cedex - France
          Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
          1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
          end enriched, double-strand cDNA was digested with Not I and cloned
          into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
          was normalized. Library was constructed by Life Technologies, a
          division of Invitrogen. This sequence belongs to sequence cluster
          2459.f
          For more information about this cluster, see
          http://www.genoscope.cns.fr/cdna?s=CS0D1018YD04Q2P1&c=2459.f.

FEATURES             source
    source
    1..1057
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="CS0D1018YD04"
    /tissue_type="PLACENTA COT 25-NORMALIZED"
    /note="1st strand cDNA was primed with a NotI-oligo(dT)
    primer. Five prime end enriched, double-strand cDNA was
    digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match      24.5%; Score 608; DB 5; Length 1057;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      720 TGTAAACCTGGGTGAGTCTTTTGGCAATGCTGCTCTATCTCTATCTCTATCTATGCT 779
Db      1 TGTAAACCTGGGTGAGTCTTTTGGCAATGCTGCTCTATCTCTATCTCTATCTATGCT 60
QY      780 CTCTGCTTGGGTGGTGTATGTTATTCATCAATCTTATCCACGCGATGTTTGTGT 839
Db      61 CTCTGCTTGGGTGGTGTATGTTATTCATCAATCTTATCCACGCGATGTTTGTGT 120
QY      840 GTTACTGATCAGAGATACGCAAGAGCTCTACATAGCATATAGCACTTCTACATGCT 899
Db      121 GTTACTGATCAGAGATACGCAAGAGCTCTACATAGCATATAGCACTTCTACATGCT 180
QY      900 GGGTTTAAATATTCATGTCAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTGA 959

```

```

Db      181 GGGTTTAAATATTCATGTCAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTGA 240
QY      960 ACACATGGCAGCTGCAGGTGCTTTTGCATTCGCTCAAGCTTATGCTTTCTTGCAATCT 1019
Db      241 ACACATGGCAGCTGCAGGTGCTTTTGCATTCGCTCAAGCTTATGCTTTCTTGCAATCT 300
QY      1020 GAGAGACCGATTAAACAAACAGAGTTCAGACCCCTTTCTTTTGGGTGTATCAGTACG 1079
Db      301 GAGAGACCGATTAAACAAACAGAGTTCAGACCCCTTTCTTTTGGGTGTATCAGTACG 360
QY      1080 TGCAGGTGCTGTGTTCCCTTAGTGTCTATCTATTGACTTATACAGCTTACATTGCACCATG 1139
Db      361 TGCAGGTGCTGTGTTCCCTTAGTGTCTATCTATTGACTTATACAGCTTACATTGCACCATG 420
QY      1140 GAGTGCAGGTTTTATTTCATTTGGGATCATCTGGGTATGCAAAAATACACATTCCAATTAT 1199
Db      421 GAGTGCAGGTTTTATTTCATTTGGGATCATCTGGGTATGCAAAAATACACATTCCAATTAT 480
QY      1200 TGCATCAGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTTTGTATCTACATAT 1259
Db      481 TGCATCAGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTTTGTATCTACATAT 540
QY      1260 TCTTGATGATACCTTCCAGCAGGCTTTGGTTCGTCATCAAAAATATCAACGATGAAG 1319
Db      541 TCTTGATGATACCTTCCAGCAGGCTTTGGTTCGTCATCAAAAATATCAACGATGAAG 600
QY      1320 AGTATTTG 1327
Db      601 AGTATTTG 608

RESULT 4
LOCUS   CR620015 1074 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0D1018YD04 of Placenta COT 25-normalized
            of Homo sapiens (human).
ACCESSION CR620015
VERSION   CR620015.1 GI:50500822
KEYWORDS HTC; CNSLT cDNA.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS   Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished
REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
          http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
          Faraday Avenue
          Genoscope.
          2 (bases 1 to 1074)
          Direct Submission
          Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
          BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
          1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
          end enriched, double-strand cDNA was digested with Not I and cloned
          into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
          was normalized. Library was constructed by Life Technologies, a
          division of Invitrogen.
          Location/Qualifiers
          1..1074
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="CS0D1018YD04"
          /tissue_type="Placenta COT 25-normalized"
          /plasmid="pCMVSPORT_6"

ORIGIN
Query Match      24.5%; Score 608; DB 3; Length 1074;
Best Local Similarity 100.0%; Pred. No. 0;

```

1691 GTCCAGGTGATGCTCTGGCTCATACATATCATGATGCGCAGGAAATATCTTAGATGATT 1750
 Db GTCCAGGTGATGCTCTGGCTCATACATATCATGATGCGCAGGAAATATCTTAGATGATT 120
 1751 TTAGAGAGCTTACTTTTGGCTAGGCAAAATACAGATGACATGACAGGAAATGCTT 1810
 Db TTAGAGAGCTTACTTTTGGCTAGGCAAAATACAGATGACATGACAGGAAATGCTT 180
 1811 GGTGGATATGGTATCAGATAGCTGGAATGCTAATAGAACTAGCTTGGTGGATAATA 1870
 Db GGTGGATATGGTATCAGATAGCTGGAATGCTAATAGAACTAGCTTGGTGGATAATA 240
 1871 ACACCTGGAATAACAGCCACATAGCTGCTGGGAAAGCTATGCTTCTAATGAACAG 1930
 Db ACACCTGGAATAACAGCCACATAGCTGCTGGGAAAGCTATGCTTCTAATGAACAG 300
 1931 CAGCCTATAAATCATGAGGACTCTAGATAGATATGTTTGGTATATTTTGGAGGG 1990
 Db CAGCCTATAAATCATGAGGACTCTAGATAGATATGTTTGGTATATTTTGGAGGG 360
 1991 TTATTTGGCTATCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAG 2050
 Db TTATTTGGCTATCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAG 420
 2051 GAGAACATCCCAAGACATTCGGGAAAGTGAATATTTTACCCACAGGAGAAATCCGGT 2110
 Db GAGAACATCCCAAGACATTCGGGAAAGTGAATATTTTACCCACAGGAGAAATCCGGT 480
 2111 TAGACAAAGCAGGATCCCTACTTTTGTGAATGGCTTATGATATAAAATGTCTACTACA 2170
 Db TAGACAAAGCAGGATCCCTACTTTTGTGAATGGCTTATGATATAAAATGTCTACTACA 540
 2171 GATTGGAGAAATCAGCTGATATTTTGGTACACCCCGAGTTTGGACCGAAACAGCTAATG 2230
 Db GATTGGAGAAATCAGCTGATATTTTGGTACACCCCGAGTTTGGACCGAAACAGCTAATG 600
 2231 CTGAGATTGGAATAAAGACATTAATTAATTTGAAGAGCTTTTACATCAGAAC 2290
 Db CTGAGATTGGAATAAAGACATTAATTAATTTGAAGAGCTTTTACATCAGAAC 660
 2291 ACTGGTTGTAGATATATAAGTAAAGACCTGATACAGGAGAGACATTAGATCACA 2350
 Db ACTGGTTGTAGATATATAAGTAAAGACCTGATACAGGAGAGACATTAGATCACA 720
 2351 AACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGAGCTACCAAAA 2410
 Db AACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGAGCTACCAAAA 780
 2411 GGAAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAAAGCAAGAAATATCTAAGA 2470
 Db GGAAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAAAGCAAGAAATATCTAAGA 840
 2471 AGACTGTTTAA 2481
 Db AGACTGTTTAA 851

RESULT 2
 CF995035
 LOCUS
 DEFINITION AGENCOURT 15622198 NIH MGC 147 Homo sapiens cdna clone
 IMAGE:30528919 5', mRNA sequence.
 ACCESSION CF995035
 VERSION CF995035.1 GI:38511095
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 861)
 NIH-MGC <http://mgi.nci.nih.gov/>
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE

JOURNAL
COMMENT

Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: gcapbs@mail.nih.gov
 Tissue procurement: Dr. Stefan Hansson
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
 and advice from Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: MDAMB15 row: k column: 08
 High quality sequence stop: 625.

FEATURES
source

1. 861
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30528919"
 /tissue_type="Human Placenta"
 /lab_host="DH10B Tona"
 /clone_lib="NIH MGC 147"
 /notes="Organ: placenta; Vector: pBluescriptR; Site 1:
 all-XhoI; Site 2: BamH; Oligo-dT primed using primer
 5'-TTTTTTTTTTTTTTVN-3', size-selected for average
 insert size 2.3 kb and normalized to 50% This is a
 primary library enriched for full-length clones and
 constructed using the Cap-trapper method (Carninci, in
 preparation). Library constructed by M. Brownstein
 (NIH/NHGRI, National Institutes of Health). Note: This is
 a NIH MGC library."

ORIGIN

Query Match 24.9%; Score 618; DB 7; Length 861;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 618; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1681 GCCTACTGTAGTCCCAAGTGTAGTCTCTGGCTCATACAAATCATGATGCGCAGGAAATATC 1740
 Db 27 GCCTACTGTAGTCCCAAGTGTAGTCTCTGGCTCATACAAATCATGATGCGCAGGAAATATC 86
 QY 1741 TTAGATGATTTTAGAGAGCTTACTTTTGGCTAAGCAAAATACAGATGAACATGCACGA 1800
 Db 87 TTAGATGATTTTAGAGAGCTTACTTTTGGCTAAGCAAAATACAGATGAACATGCACGA 146
 QY 1801 GTAAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG 1860
 Db 147 GTAAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG 206
 QY 1861 GTGGATAATAACACCTGGAATTAACCCACATAGCAGCTGGTGGGAAAGCTATGCTTCT 1920
 Db 207 GTGGATAATAACACCTGGAATTAACCCACATAGCAGCTGGTGGGAAAGCTATGCTTCT 266
 QY 1921 AATGAAACAGCAGCTTATAAATCATAGGACTCTAGATAGTATGTTTGGTATTT 1980
 Db 267 AATGAAACAGCAGCTTATAAATCATAGGACTCTAGATAGTATGTTTGGTATTT 326
 QY 1981 TTTGGAGGGGTTATTTGGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGG 2040
 Db 327 TTTGGAGGGGTTATTTGGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGG 386
 QY 2041 ATAGCTGAAGAGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAGGGA 2100
 Db 387 ATAGCTGAAGAGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAGGGA 446
 QY 2101 GAATTCCTGTTAGACAAACAGCAGTCCCTACTTTTGGTGAATTCGCTTATGTATAAATG 2160
 Db 447 GAATTCCTGTTAGACAAACAGCAGTCCCTACTTTTGGTGAATTCGCTTATGTATAAATG 506
 QY 2161 TCATATCAGATTTTGGAGAAATGACGTGGATTTTTCGTACACCCCAAGGTTTGGACCGA 2220

Dp 181 GACTCTAGATGATGATTATGTTTGGTTATTTTGGAGGGTTATTGGCTATTCTCGTGA 240
Qy 2010 TGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAGGAGACATCCCAAGACAT 2069
Dp 241 TGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAGGAGACATCCCAAGACAT 300
Qy 2070 TCGGAAAGTGACTATTTTACCCACAGGAGAAATCCGTGTAGACAAAGCAGGATCCCC 2129
Dp 301 TCGGAAAGTGACTATTTTACCCACAGGAGAAATCCGTGTAGACAAAGCAGGATCCCC 360
Qy 2130 TACTTTGTGAATGCTTATGTATATAAATGTCATACAGATTGGAGAAATGCAGCT 2189
Dp 361 TACTTTGTGAATGCTTATGTATATAAATGTCATACAGATTGGAGAAATGCAGCT 420
Qy 2190 GGATTTTCGT 2199
Dp 421 GGATTTTCGT 430

Search completed: December 16, 2004, 16:17:34
Job time : 2198.32 secs

```
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-878-178-717

Query Match      17.3%; Score 430; DB 9; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.4e-216;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1770 GCTAAGCGAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 1829
Db 1 GCTAAGCGAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 60

QY 1830 GATAGCTGGTGGGAAAGCTATGCTCTTAATGAAGAGAGATTAACACCTGGAATAACAGCCA 1889
Db 61 GATAGCTGGTGGGAAAGCTATGCTCTTAATGAAGAGAGATTAACACCTGGAATAACAGCCA 120

QY 1890 CATAGCACTGGTGGGAAAGCTATGCTCTTAATGAAGAGAGATTAACACCTGGAATAACAGCCA 1949
Db 121 CATAGCACTGGTGGGAAAGCTATGCTCTTAATGAAGAGAGATTAACACCTGGAATAACAGCCA 180

QY 1950 GACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGGTGA 2009
Db 181 GACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGGTGA 240

QY 2010 TGATATCAACAAATTTCTCTGGATGTTTAGGATAGCTGAAGGAGAAACATCCCAAAGACAT 2069
Db 241 TGATATCAACAAATTTCTCTGGATGTTTAGGATAGCTGAAGGAGAAACATCCCAAAGACAT 300

QY 2070 TCGGGAAGTGAATTTTACCCACAGGAGAGATTTCCGCTAGACAAAGCAGGATCCCC 2129
Db 301 TCGGGAAGTGAATTTTACCCACAGGAGAGATTTCCGCTAGACAAAGCAGGATCCCC 360

QY 2130 TACTTTGTTGAATGCTCTATGTAATAAATGTCATACAGATTTGGAGAAATGCAGCT 2189
Db 361 TACTTTGTTGAATGCTCTATGTAATAAATGTCATACAGATTTGGAGAAATGCAGCT 420

QY 2190 GGATTTTCGT 2199
Db 421 GGATTTTCGT 430

RESULT 15
US-10-146-502-717
; Sequence 717, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146.502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-146-502-717

Query Match      17.3%; Score 430; DB 14; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.4e-216;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1770 GCTAAGCGAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 1829
Db 1 GCTAAGCGAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 60

QY 1830 GATAGCTGGTGGGAAAGCTATGCTCTTAATGAAGAGAGATTAACACCTGGAATAACAGCCA 1889
Db 61 GATAGCTGGTGGGAAAGCTATGCTCTTAATGAAGAGAGATTAACACCTGGAATAACAGCCA 120

QY 1890 CATAGCACTGGTGGGAAAGCTATGCTCTTAATGAAGAGAGATTAACACCTGGAATAACAGCCA 1949
Db 121 CATAGCACTGGTGGGAAAGCTATGCTCTTAATGAAGAGAGATTAACACCTGGAATAACAGCCA 180

QY 1950 GACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGGTGA 2009
Db 181 GACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGGTGA 240

QY 2010 TGATATCAACAAATTTCTCTGGATGTTTAGGATAGCTGAAGGAGAAACATCCCAAAGACAT 2069
Db 241 TGATATCAACAAATTTCTCTGGATGTTTAGGATAGCTGAAGGAGAAACATCCCAAAGACAT 300

QY 2070 TCGGGAAGTGAATTTTACCCACAGGAGAGATTTCCGCTAGACAAAGCAGGATCCCC 2129
Db 301 TCGGGAAGTGAATTTTACCCACAGGAGAGATTTCCGCTAGACAAAGCAGGATCCCC 360

QY 2130 TACTTTGTTGAATGCTCTATGTAATAAATGTCATACAGATTTGGAGAAATGCAGCT 2189
Db 361 TACTTTGTTGAATGCTCTATGTAATAAATGTCATACAGATTTGGAGAAATGCAGCT 420

QY 2190 GGATTTTCGT 2199
Db 421 GGATTTTCGT 430

RESULT 14
US-10-046-935-717
; Sequence 717, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046.935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-046-935-717

Query Match      17.3%; Score 430; DB 13; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.4e-216;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1770 GCTAAGCGAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 1829
```

RESULT 11

US-10-296-115-629
; Sequence 629, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 629
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-629

Query Match 26.3%; Score 653; DB 16; Length 1114;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 653; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 177 CGGGGGCTGTGCGAGCCGGCTGGGTGGAGTGGTCTCTCTCCATCACCATCCTCTCTCT 236
Db 388 CGGGGGCTGTGCGAGCCGGCTGGGTGGAGTGGTCTCTCTCTCTCTCTCTCTCTCTCT 447
QY 237 GGCCTGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGTCATCCGCTTCGAAGCATCAT 296
Db 448 GGCCTGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGTCATCCGCTTCGAAGCATCAT 507
QY 297 CACGAGTTCGACCCGCTGGTTAACTATAGATCAACATCATCTTGCATCTCATGGTT 356
Db 508 CCACGAGTTCGACCCGCTGGTTAACTATAGATCAACATCATCTTGCATCTCATGGTT 567
QY 357 CTATGAATTTTAAATGGTTTGAATGAAGACATGGTATCCATAGAAAGATAGG 416
Db 568 CTATGAATTTTAAATGGTTTGAATGAAGACATGGTATCCATAGAAAGATAGG 627
QY 417 TGGTACTGTTTACCAGGGTTCATGATTAACCCGCTGGCTTATTCATGATTTTAAATAC 476
Db 628 TGGTACTGTTTACCAGGGTTCATGATTAACCCGCTGGCTTATTCATGATTTTAAATAC 687
QY 477 ATTGAACATACTGTTCATAGACATAGACATAGTGTGTTCCTTGACCAACTTTTAGCGG 536
Db 688 ATTGAACATACTGTTCATAGACATAGACATAGTGTGTTCCTTGACCAACTTTTAGCGG 747
QY 537 CTTTACATCTATATCTACTTTCCTGCTTACAGAGAACTTTTGGACCAAGCAGGACT 596
Db 748 CTTTACATCTATATCTACTTTCCTGCTTACAGAGAACTTTTGGACCAAGCAGGACT 807
QY 597 TTTAGCTGCTTTTATGCTATTTGTACAGGCTACATATCTCGGTCAGTGGATC 656
Db 808 TTTAGCTGCTTTTATGCTATTTGTACAGGCTACATATCTCGGTCAGTGGATC 867
QY 657 CTTTCATATGAAGCATTTGCTATTTTGGCACTTCAGTTCATATCTATTTTATGGTAAA 716
Db 868 CTTTCATATGAAGCATTTGCTATTTTGGCACTTCAGTTCATATCTATTTTATGGTAAA 927
QY 717 ATCTGTAATACTGGTTCAGTTTGTGGACAAATGTGCTGCTTATCTATTTCTATAT 776
Db 928 ATCTGTAATACTGGTTCAGTTTGTGGACAAATGTGCTGCTTATCTATTTCTATAT 987
QY 777 GGTCTCTCTGGGTGGTATGATTTATCATCAATCTTATTCACCTGCATG 829
Db 988 GGTCTCTCTGGGTGGTATGATTTATCATCAATCTTATTCACCTGCATG 1040

RESULT 12

US-09-998-598-1643
; Sequence 1643, Application US/09998598

Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Mesgher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1643
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1643

Query Match 20.2%; Score 500; DB 9; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.1e-253;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 471 AAATACATTGAACATACTCTTACATAGAGAGTATGTGTCTTGCACCACTTT 530
Db 1 AAATACATTGAACATACTCTTACATAGAGAGTATGTGTCTTGCACCACTTT 60
QY 531 TAGCGGCTTACATCTATATCTACTTCTCTCTTACAAGAGAACTTTGGAACCAAGGAGC 590
Db 61 TAGCGGCTTACATCTATATCTACTTCTCTCTTACAAGAGAACTTTGGAACCAAGGAGC 120
QY 591 AGGACTTTTACGCTGCTTTTATTTGCTATTTGACAGGCTACATATCTCGGTCACTAGC 650
Db 121 AGGACTTTTACGCTGCTTTTATTTGCTATTTGACAGGCTACATATCTCGGTCACTAGC 180
QY 651 TGGATCTTTGATTAATGAAGCATTCCTATTTTGGCACTTCAGTTCACATATCTATTTATG 710
Db 181 TGGATCTTTGATTAATGAAGCATTCCTATTTTGGCACTTCAGTTCACATATCTATTTATG 240
QY 711 GGTAAATCTGTAATAAATCTGGTCACTTTTGGCAAACTGCTGCTTATCTCTATTT 770
Db 241 GGTAAATCTGTAATAAATCTGGTCACTTTTGGCAAACTGCTGCTTATCTCTATTT 300
QY 771 CTATATGCTCTGCTTGGGTGGTATGATTTATCATCAATCTTATTTCCATGCTATG 830
Db 301 CTATATGCTCTGCTTGGGTGGTATGATTTATCATCAATCTTATTTCCATGCTATG 360
QY 831 ATTGTGTTGTTACTGATGAGAGATACAGCAAGAGTCTACATAGCATATAGCACTTT 890
Db 361 ATTGTGTTGTTACTGATGAGAGATACAGCAAGAGTCTACATAGCATATAGCACTTT 420
QY 891 CTACATTTGGGTTTAAATTTATTTATCATGAGATACCTTTTGGGATTCAGCAATCAG 950
Db 421 CTACATTTGGGTTTAAATTTATTTATCATGAGATACCTTTTGGGATTCAGCAATCAG 480
QY 951 AACAACTGACACATGGCAG 970
Db 481 AACAACTGACACATGGCAG 500

RESULT 13

US-09-878-178-717
; Sequence 717, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08

QY 267 CTTGCGCGTCATCCGCTTCGAAAGCATCCACGAGTTCGACCGTGGTTAACTATAG 326
Db 1 CTTGCGCGTCATCCGCTTCGAAAGCATCCACGAGTTCGACCGTGGTTAACTATAG 60
QY 327 ATCAACACATCATCTGCTCATCTCATGGTCTTATGAATTTTAAATTTGGTTGATGAAG 386
Db 61 ATCAACACATCATCTGCTCATCTCATGGTCTTATGAATTTTAAATTTGGTTGATGAAG 120
QY 387 AGCATGTTATCCACTAGGAGAAATAGTAGTGCTACTGTTTACCAGCGGTTGATGAATAC 446
Db 121 AGCATGTTATCCACTAGGAGAAATAGTAGTGCTACTGTTTACCAGCGGTTGATGAATAC 180
QY 447 CGTGGCGTTATTCATGGAATTTTAAATATGATGAATTAATGATGAATGATGAAGCGT 506
Db 181 CGTGGCGTTATTCATGGAATTTTAAATATGATGAATTAATGATGAATGATGAAGCGT 240
QY 507 ATGTGTTGCTTTCGACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 566
Db 241 ATGTGTTGCTTTCGACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 300
QY 567 AAGAGACTTTGGAACCAAGGACGAGCTTTTAGCTGCTGCTTTTATGCTATGTTAC 626
Db 301 AAGAGACTTTGGAACCAAGGACGAGCTTTTAGCTGCTGCTTTTATGCTATGTTAC 360
QY 627 AGCTACATATCTCGTTCAGTAGCTGATCTCTTGTATGATGAAGCAATGCTATTTTTCG 686
Db 361 AGCTACATATCTCGTTCAGTAGCTGATCTCTTGTATGATGAAGCAATGCTATTTTTCG 420
QY 687 ACTTCAGTTCACATATCTATTTTGGTAAATCTGTAATGAGTGGTCACTTTTGGTAC 746
Db 421 ACTTCAGTTCACATATCTATTTTGGTAAATCTGTAATGAGTGGTCACTTTTGGTAC 480
QY 747 AATGTGCTGCTTATCTATTTCTATATGTTCTGCTTGGGTTGTTATGTTAT 806
Db 481 AATGTGCTGCTTATCTATTTCTATATGTTCTGCTTGGGTTGTTATGTTAT 540
QY 807 CATCAATCTTATTCCTGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 866
Db 541 CATCAATCTTATTCCTGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 600
QY 867 AGTCTACATAGCATATAGCACTTCTACATTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 926
Db 601 AGTCTACATAGCATATAGCACTTCTACATTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 660
QY 927 TTTTGTGGGATTCAGCCATTCAGCAAGTGAACATGCGAGTGGAGTGGTCTTTTCG 986
Db 661 TTTTGTGGGATTCAGCCATTCAGCAAGTGAACATGCGAGTGGAGTGGTCTTTTCG 720
QY 987 ATTGCTGCAAGCTTATGCTTTCTTTCAGTATCTGAGAGACCGATTAAACAAACAGAGTT 1046
Db 721 ATTGCTGCAAGCTTATGCTTTCTTTCAGTATCTGAGAGACCGATTAAACAAACAGAGTT 780
QY 1047 CCAGACCTTTTCTTTTGGGTTATCTAGTGCAGTGTGTTGTTGTTGTTGTTGTTGTTGTT 1106
Db 781 CCAGACCTTTTCTTTTGGGTTATCTAGTGCAGTGTGTTGTTGTTGTTGTTGTTGTTGTT 840
QY 1107 CTATTTGACATTACAGGTTTACATTCACCATGGAGTGGCAGGTTTATTTATTTGTTGTTGTT 1166
Db 841 CTATTTGACATTACAGGTTTACATTCACCATGGAGTGGCAGGTTTATTTATTTGTTGTTGTT 900
QY 1167 TACTGGGTATGCAAAATACACATTCGAATTTATGATGATGATGATGATGATGATGATGATGAT 1226
Db 901 TACTGGGTATGCAAAATACACATTCGAATTTATGATGATGATGATGATGATGATGATGATGAT 960
QY 1227 GACTTGGGTGTTCTTTCTTTCTTTGATCTACATATTTCTTTGATGATGATGATGATGATGATGAT 1286
Db 961 GACTTGGGTGTTCTTTCTTTCTTTGATCTACATATTTCTTTGATGATGATGATGATGATGATGAT 1020
QY 1287 TTGTTTCTGATCAAAATATCAAGATGAAGA 1320
Db 1021 TTGTTTCTGATCAAAATATCAAGATGAAGA 1054

RESULT 9
US-10-621-401-133
; Sequence 133, Application US/10621401
; Publication No. US2004003827A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P202P2C1
; CURRENT APPLICATION NUMBER: US/10/621,401
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 09/974,879
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/239,893
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/818,683
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 09/305,736
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: PCT/US98/23435
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/064,911
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,912
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,983
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,900
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,988
; PRIOR FILING DATE: 1997-11-07
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-621-401-133

Query Match 42.5%; Score 1054; DB 16; Length 1543;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 267 CTTGCGCGTCATCCGCTTCGAAAGCATCCACGAGTTCGACCGTGGTTAACTATAG 326
Db 1 CTTGCGCGTCATCCGCTTCGAAAGCATCCACGAGTTCGACCGTGGTTAACTATAG 60
QY 327 ATCAACACATCATCTGCTCATCTCATGGTCTTATGAATTTTAAATTTGGTTGATGAAG 386
Db 61 ATCAACACATCATCTGCTCATCTCATGGTCTTATGAATTTTAAATTTGGTTGATGAAG 120
QY 387 AGCATGTTATCCACTAGGAGAAATAGTAGTGCTACTGTTTACCAGCGGTTGATGAATAC 446
Db 121 AGCATGTTATCCACTAGGAGAAATAGTAGTGCTACTGTTTACCAGCGGTTGATGAATAC 180
QY 447 CGTGGCGTTATTCATGGAATTTTAAATATGATGAATTAATGATGAATGATGAAGCGT 506
Db 181 CGTGGCGTTATTCATGGAATTTTAAATATGATGAATTAATGATGAATGATGAAGCGT 240
QY 507 ATGTGTTGCTTTCGACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 566
Db 241 ATGTGTTGCTTTCGACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 300
QY 567 AAGAGAACTTTGGAACCAAGGACGAGCTTTTAGCTGCTGCTTTTATGCTATGTTAC 626

Db 721 ATTGCTGCAAGCTTATGCTTTCTTTGAGTATCTGAGACCGATTAAACAAACAGAGTT 780
QY 1047 CAGAGCCCTTTCTTTTGGGTGATACATAGTGTGCTGTGTTCTTGTAGTGTAT 1106
Db 781 CAGAGCCCTTTCTTTTGGGTGATACATAGTGTGCTGTGTTCTTGTAGTGTAT 840
QY 1107 CTATTTGACTATACAGGTTACATTCACCATGGAGTGGCAGGTTTATTTCATTGTGGGA 1166
Db 841 CTATTTGACTATACAGGTTACATTCACCATGGAGTGGCAGGTTTATTTCATTGTGGGA 900
QY 1167 TACTGGGTATGCAAAATACAGTTCCCAATTATTGATGCTGAGCATCAACCTAC 1226
Db 901 TACTGGGTATGCAAAATACAGTTCCCAATTATTGATGCTGAGCATCAACCTAC 960
QY 1227 GACTTGGGTGCTTTCTTTGATCTACATATTTCTTGTATGTAACCTTCCAGCAGGCT 1286
Db 961 GACTTGGGTGCTTTCTTTGATCTACATATTTCTTGTATGTAACCTTCCAGCAGGCT 1020
QY 1287 TTGGTTCTGATCAAAATATCAAGATGAAAGA 1320
Db 1021 TTGGTTCTGATCAAAATATCAAGATGAAAGA 1054

RESULT 6

US-09-305-736-133
; Sequence 133, Application US/09305736
; Publication No. US2003008078A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/305,736
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: PCT/US98/23435
; EARLIER FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/064,911
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,912
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,983
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,900
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,988
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,987
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,908
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,984
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,985
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/066,094
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,100
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,089
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,095
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,090
; EARLIER FILING DATE: 1997-11-17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)

; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-305-736-133

Query Match 42.5%; Score 1054; DB 10; Length 1543;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 267 CTTGCGCGTATCCGCTTCCGAAAGCATCATCCAGAGTTCGACCCGCTGGTTAACTATAG 326
Db 1 CTTGCGCGTATCCGCTTCCGAAAGCATCATCCAGAGTTCGACCCGCTGGTTAACTATAG 60
QY 327 ATCAACACATCATCTTCCATCTCATGGTCTCTATGAAATTTTAAATTTGGTTGATGAAAG 386
Db 61 ATCAACACATCATCTTCCATCTCATGGTCTCTATGAAATTTTAAATTTGGTTGATGAAAG 120
QY 387 AGCATGTATCCACTAGGAAGAATAGTAGTGTGTACTGTGTATCCAGGGTGTGATGATAAC 446
Db 121 AGCATGTATCCACTAGGAAGAATAGTAGTGTGTACTGTGTATCCAGGGTGTGATGATAAC 180
QY 447 CGTGGCCTTATTCATTTGGATTTTAAATACATTTGAACATTAACCTGTTCAATAGACGT 506
Db 181 CGTGGCCTTATTCATTTGGATTTTAAATACATTTGAACATTAACCTGTTCAATAGACGT 240
QY 507 ATGTGTGTTCTTGCACCAACTTTTAGCGGCTTACATCTATATCTACTTTCTGCTTAC 566
Db 241 ATGTGTGTTCTTGCACCAACTTTTAGCGGCTTACATCTATATCTACTTTCTGCTTAC 300
QY 567 AAGAGAACTTTGGAACCAAGGACGAGGACTTTAGTGTGTTTATTTGCTATTGTACC 626
Db 301 AAGAGAACTTTGGAACCAAGGACGAGGACTTTAGTGTGTTTATTTGCTATTGTACC 360
QY 627 AGGTACATATCTCGGTGCTAGTGTGATGCTTCTGTATGTAATGAAGGCAATTTGCTTGC 686
Db 361 AGGTACATATCTCGGTGCTAGTGTGATGCTTCTGTATGTAATGAAGGCAATTTGCTTGC 420
QY 687 ACTTCAGTTCACTACTATTTATTTGGTGAATCTGTAAACCTGGGTGAGTTTGTGGAC 746
Db 421 ACTTCAGTTCACTACTATTTATTTGGTGAATCTGTAAACCTGGGTGAGTTTGTGGAC 480
QY 747 AATGTGCTGCTGTATCTCTATTTCTATATGCTCTGCTTGGGTGCTTATCTATTAT 806
Db 481 AATGTGCTGCTGTATCTCTATTTCTATATGCTCTGCTTGGGTGCTTATCTATTAT 540
QY 807 CATCAATCTTATCCACTGATCTATTTGTTGTTACTGATGATGATGATGATGATGATGATGAT 866
Db 541 CATCAATCTTATCCACTGATCTATTTGTTGTTACTGATGATGATGATGATGATGATGATGAT 600
QY 867 AGTCTACATAGCATATAGCACTTTCTACATTTGGGTGTTAAATATTAATCAATGATGATGAT 926
Db 601 AGTCTACATAGCATATAGCACTTTCTACATTTGGGTGTTAAATATTAATCAATGATGATGAT 660
QY 927 TTTTGTGGGATTCAGCCCAATCAGAAAGTGAACACATGGCAGCTGCGAGGTGCTTTGC 986
Db 661 TTTTGTGGGATTCAGCCCAATCAGAAAGTGAACACATGGCAGCTGCGAGGTGCTTTGC 720
QY 987 ATTGCTGCAAGCTTATGCTTTCTTGTGAGTATCTGAGAGCCGATTTAAACAAACAGAGTT 1046
Db 721 ATTGCTGCAAGCTTATGCTTTCTTGTGAGTATCTGAGAGCCGATTTAAACAAACAGAGTT 780
QY 1047 CCAGACCCCTTTCTTTTGGGTGATCACTAGTGTGCTGCTGCTTCTTAGTGTCTAT 1106
Db 781 CCAGACCCCTTTCTTTTGGGTGATCACTAGTGTGCTGCTGCTTCTTAGTGTCTAT 840
QY 1107 CTATTTGACTTATACAGGTTTACATTCACCATGGAGTGGCAGGTTTATTTCATTGTGGGA 1166
Db 841 CTATTTGACTTATACAGGTTTACATTCACCATGGAGTGGCAGGTTTATTTCATTGTGGGA 900
QY 1167 TACTGGGTATGCAAAATACACATTTCAATTTATGATGATGATGATGATGATGATGATGATGAT 1226

Db 601 AGCTACATAGCATATAGCACCTTTCTACATTTGTGGTTTAAATATATCAATGCAGATACC 660
Qy 927 TTTTGTGGATCCAGCAATCAGAACAGTGAACACATGCGAGTGCAGGTGCTTTGC 986
Db 661 TTTTGTGGATCCAGCAATCAGAACAGTGAACACATGCGAGTGCAGGTGCTTTGC 720
Qy 987 ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGACCCGATTAACAAACAAGAGTT 1046
Db 721 ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGACCCGATTAACAAACAAGAGTT 780
Qy 1047 CCAGACCCCTTTCTTTTGGGTGATCACTAGCTCAGGTGCTGTTCTTGTAGTGCAT 1106
Db 781 CCAGACCCCTTTCTTTTGGGTGATCACTAGCTCAGGTGCTGTTCTTGTAGTGCAT 840
Qy 1107 CTATTGACTTATACAGGTTCATTTGCACCATGAGTGCAGGTTTATTCATTGTGGGA 1166
Db 841 CTATTGACTTATACAGGTTCATTTGCACCATGAGTGCAGGTTTATTCATTGTGGGA 900
Qy 1167 TACTGGGTATGCAAAATATACATTCCTCAATTTATTCATCAGTGTGAGCATCAACCTAC 1226
Db 901 TACTGGGTATGCAAAATATACATTCCTCAATTTATTCATCAGTGTGAGCATCAACCTAC 960
Qy 1227 GACTTGGGTGCTTTCTTTTGTGATCTACATATCTTGTATGTAACCTTCCAGCAGGCT 1286
Db 961 GACTTGGGTGCTTTCTTTTGTGATCTACATATCTTGTATGTAACCTTCCAGCAGGCT 1020
Qy 1287 TTGGTTCTGCATCAAAATATCAACGATGAAGA 1320
Db 1021 TTGGTTCTGCATCAAAATATCAACGATGAAGA 1054

RESULT 5

US-09-974-879-133
; Sequence 133, Application US/09974879
; Publication No. US2003028003A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P2
; CURRENT APPLICATION NUMBER: US/09/974,879
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/239,893
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/818,683
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 09/305,736
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: PCT/US98/23435
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/064,911
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,912
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,983
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,900
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,988
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,987
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,908
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,984
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,985
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/066,094
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,100
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,089

Query Match 42.5%; Score 1054; DB 10; Length 1543;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 267 CTTTCGGCGTTCATCGGCTTCGAAAGCATCATCCAGAGTTCGACCCGCTGTTAACTATAG 326
Db 1 CTTTCGGCGTTCATCGGCTTCGAAAGCATCATCCAGAGTTCGACCCGCTGTTAACTATAG 60
Qy 327 ATCAACACATCATCTTTCGATCTCATGGGTTCTATGAATTTTAAATTTGGTTGATGAAG 386
Db 61 ATCAACACATCATCTTTCGATCTCATGGGTTCTATGAATTTTAAATTTGGTTGATGAAG 120
Qy 387 AGCATGSTATCCACTAGGAAGATAGTAGTGTGTACTGTTTACCCAGGGTTTCATGATAAC 446
Db 121 AGCATGSTATCCACTAGGAAGATAGTAGTGTGTACTGTTTACCCAGGGTTTCATGATAAC 180
Qy 447 CGCTGGCCTTATTCATTGGATTTTAAATACATTTGAACATACTGTTTCATGAAGACGT 506
Db 181 CGCTGGCCTTATTCATTGGATTTTAAATACATTTGAACATACTGTTTCATGAAGACGT 240
Qy 507 ATGTGTGTTCTTGCACCAACTTTTAGCGGCTTACATCTATATCTACTTTCTCTGTTAC 566
Db 241 ATGTGTGTTCTTGCACCAACTTTTAGCGGCTTACATCTATATCTACTTTCTCTGTTAC 300
Qy 567 AAGAGAACTTTTGAACCAAGAGGAGGAGTCTTTAGCTGCTGTTGTTTATTCATTTGATCC 626
Db 301 AAGAGAACTTTTGAACCAAGAGGAGGAGTCTTTAGCTGCTGTTGTTTATTCATTTGATCC 360
Qy 627 AGGCTACATATCTCGGTGAGTAGTGTGATCTTGTGATATGAAGGATTCATTTTTCG 686
Db 361 AGGCTACATATCTCGGTGAGTAGTGTGATCTTGTGATATGAAGGATTCATTTTTCG 420
Qy 687 ACTTCAGTTCACATATCTATTATGGGTGATAAATCTGTAATAAATCTGGGTGAGTTTGGAC 746
Db 421 ACTTCAGTTCACATATCTATTATGGGTGATAAATCTGTAATAAATCTGGGTGAGTTTGGAC 480
Qy 747 AATGTGCTGCTGCTTATCTTCTTATATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 806
Db 481 AATGTGCTGCTGCTTATCTTCTTATATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Qy 807 CATCAATCTTATTCCTGATGATTTTGTGTTGTTTACTGATGACAGATACAGCAAAAG 866
Db 541 CATCAATCTTATTCCTGATGATTTTGTGTTGTTTACTGATGACAGATACAGCAAAAG 600
Qy 867 AGTCTACATAGCATATAGCACTTTCTACATTTGGGTTTAAATATTAATATCAATGCAGATCC 926
Db 601 AGTCTACATAGCATATAGCACTTTCTACATTTGGGTTTAAATATTAATATCAATGCAGATCC 660
Qy 927 TTTTGTGGGATTCAGCCCATCAGAACAGTGAACATGCGAGTGCAGGTGCTTTTTCG 986
Db 661 TTTTGTGGGATTCAGCCCATCAGAACAGTGAACATGCGAGTGCAGGTGCTTTTTCG 720
Qy 987 ATTGCTGCAAGCTTATGCTTTTCTTGTGAGTATCTGAGAGCCGATTAACAAACAAGAGTT 1046

		Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	267	CTTCGGCGTCATCGGCTCGAAAGCATCATCCACGAGTTCCGACCGGTGTTAACTATAG	326		
Db	1	CTTCGGCGTCATCGGCTCGAAAGCATCATCCACGAGTTCCGACCGGTGTTAACTATAG	60		
Qy	327	ATCAACACATCATCTTGGCATCTCATGGTTCATGAATTTTAAATTTGGTTTGATGAAG	386		
Db	61	ATCAACACATCATCTTGGCATCTCATGGTTCATGAATTTTAAATTTGGTTTGATGAAG	120		
Qy	387	AGCATGGTATCCACTAGGAAGATAGTAGTGTCTAGTTTACCCAGGGTTCATGATAAC	446		
Db	121	AGCATGGTATCCACTAGGAAGATAGTAGTGTCTAGTTTACCCAGGGTTCATGATAAC	180		
Qy	447	CGTGGCCCTTATCATATGATTTTAAATACATGAACATACTGTTTACATAAGAGAGCT	506		
Db	181	CGTGGCCCTTATCATATGATTTTAAATACATGAACATACTGTTTACATAAGAGAGCT	240		
Qy	507	ATGTGTGTTCTTGGACCACTTTTAGCGGCTTACATCTATATCTACTTTCCTGCTTAC	566		
Db	241	ATGTGTGTTCTTGGACCACTTTTAGCGGCTTACATCTATATCTACTTTCCTGCTTAC	300		
Qy	567	AGAGAACTTTGGAACCAAGGACGAGACTTTTAGCTGCTTTTATTTGCTATTGTACC	626		
Db	301	AGAGAACTTTGGAACCAAGGACGAGACTTTTAGCTGCTTTTATTTGCTATTGTACC	360		
Qy	627	AGCTACATATCTCGGTCTAGTGTGATCTGATCTATATCTACTTTCCTGCTTAC	686		
Db	361	AGCTACATATCTCGGTCTAGTGTGATCTGATCTATATCTACTTTCCTGCTTAC	420		
Qy	687	ACTTCAGTTTCACATCTATTTTATGGTAAATCTGTAAATCTGGTTCAGTTTTCGAC	746		
Db	421	ACTTCAGTTTCACATCTATTTTATGGTAAATCTGTAAATCTGGTTCAGTTTTCGAC	480		
Qy	747	ATGTGCTGCTGTTATCTTATTTCTATATGTTCTGCTGCTGCTGCTGCTGCTGCTGCT	806		
Db	481	ATGTGCTGCTGTTATCTTATTTCTATATGTTCTGCTGCTGCTGCTGCTGCTGCTGCT	540		
Qy	807	CATCAATCTTATTCACATCTGATTTTCTGTTGTTTCTGATGTCAGAGATACAGAAAG	866		
Db	541	CATCAATCTTATTCACATCTGATTTTCTGTTGTTTCTGATGTCAGAGATACAGAAAG	600		
Qy	867	AGTCTACATAGCATATAGCATTTCTACATTTGGGTTTAAATTTATCAATGCAGATACC	926		
Db	601	AGTCTACATAGCATATAGCATTTCTACATTTGGGTTTAAATTTATCAATGCAGATACC	660		
Qy	927	TTTTGGGATTTCCAGCCCAATCAGACAGTGAACACATGGCAGCTGAGGTTCTTTC	986		
Db	661	TTTTGGGATTTCCAGCCCAATCAGACAGTGAACACATGGCAGCTGAGGTTCTTTC	720		
Qy	987	ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGACCGGATTAACAAACACAGAGTT	1046		
Db	721	ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGACCGGATTAACAAACACAGAGTT	780		
Qy	1047	CCAGACCTTTCTTTTGGGTTATCACTAGCTGAGGCTGCTGCTGCTGCTGCTGCTGCT	1106		
Db	781	CCAGACCTTTCTTTTGGGTTATCACTAGCTGAGGCTGCTGCTGCTGCTGCTGCTGCT	840		
Qy	1107	CTATTGACTTATACAGGTTTACATTCACCACTGAGTGGCAGGTTTATTCATTGTGGGA	1166		
Db	841	CTATTGACTTATACAGGTTTACATTCACCACTGAGTGGCAGGTTTATTCATTGTGGGA	900		
Qy	1167	TACTGGGATGAAAAATACACATTTCCAAATTTTGCATCAGTGTCTGAGCATCAACCTTAC	1226		
Db	901	TACTGGGATGAAAAATACACATTTCCAAATTTTGCATCAGTGTCTGAGCATCAACCTTAC	960		
Qy	1227	GACTTGGGTTCTTTCTTTCTTTGATCTACATATCTTGTATGATCTTCCAGAGGCT	1286		
Db	961	GACTTGGGTTCTTTCTTTCTTTGATCTACATATCTTGTATGATCTTCCAGAGGCT	1020		
Qy	1287	TTGGTTCTGCATCAAAAATATCAACGATGAAGA	1320		
Db	1021	TTGGTTCTGCATCAAAAATATCAACGATGAAGA	1054		

RESULT 4
 US-10-264-237-412
 ; Sequence 412, Application US/10264237
 ; Publication No. US20040009491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birst et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P131P1
 ; CURRENT APPLICATION NUMBER: US/10/264,237
 ; PRIOR FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450
 ; PRIOR FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: US 60/205,515
 ; PRIOR FILING DATE: 2000-05-19
 ; NUMBER OF SEQ ID NOS: 2876
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 412
 ; LENGTH: 1209
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1203)..(1203)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-10-264-237-412

		Query Match 42.5%; Score 1054; DB 16; Length 1209; Best Local Similarity 100.0%; Pred. No. 0; Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	267	CTTCGGCGTCATCGGCTCGAAAGCATCATCCACGAGTTCCGACCGGTGTTAACTATAG	326		
Db	1	CTTCGGCGTCATCGGCTCGAAAGCATCATCCACGAGTTCCGACCGGTGTTAACTATAG	60		
Qy	327	ATCAACACATCATCTTGGCATCTCATGGTTCATGAATTTTAAATTTGGTTTGATGAAG	386		
Db	61	ATCAACACATCATCTTGGCATCTCATGGTTCATGAATTTTAAATTTGGTTTGATGAAG	120		
Qy	387	AGCATGGTATCCACTAGGAAGATAGTAGTGTGATCTGTTTACCCAGGGTTCATGATAAC	446		
Db	121	AGCATGGTATCCACTAGGAAGATAGTAGTGTGATCTGTTTACCCAGGGTTCATGATAAC	180		
Qy	447	CGTGGCCCTTATCATATGATTTTAAATACATGAACATACTGTTTACATAAGAGAGCT	506		
Db	181	CGTGGCCCTTATCATATGATTTTAAATACATGAACATACTGTTTACATAAGAGAGCT	240		
Qy	507	ATGTGTGTTCTTGCACCACTTTTAGCGGCTTACATCTATATCTACTTTCCTGCTTAC	566		
Db	241	ATGTGTGTTCTTGCACCACTTTTAGCGGCTTACATCTATATCTACTTTCCTGCTTAC	300		
Qy	567	AGAGAACTTTGGAACCAAGGACGAGACTTTTAGCTGCTTTTATTTGCTATTGTACC	626		
Db	301	AGAGAACTTTGGAACCAAGGACGAGACTTTTAGCTGCTTTTATTTGCTATTGTACC	360		
Qy	627	AGCTACATATCTCGGTCTAGTGTGATCTGATCTATATCTACTTTCCTGCTTAC	686		
Db	361	AGCTACATATCTCGGTCTAGTGTGATCTGATCTATATCTACTTTCCTGCTTAC	420		
Qy	687	ACTTCAGTTTCACATCTATTTTATGGTAAATCTGTAAATCTGGTTCAGTTTTCGAC	746		
Db	421	ACTTCAGTTTCACATCTATTTTATGGTAAATCTGTAAATCTGGTTCAGTTTTCGAC	480		
Qy	747	AATGTGCTGCTGTTATCTTATTTCTATATGTTCTGCTGCTGCTGCTGCTGCTGCTGCT	806		
Db	481	AATGTGCTGCTGTTATCTTATTTCTATATGTTCTGCTGCTGCTGCTGCTGCTGCTGCT	540		
Qy	807	CATCAATCTTATTCACATCTGATTTTCTGTTGTTTCTGATGTCAGAGATACAGAAAG	866		
Db	541	CATCAATCTTATTCACATCTGATTTTCTGTTGTTTCTGATGTCAGAGATACAGAAAG	600		
Qy	867	AGTCTACATAGCATATAGCATTTCTACATTTGGGTTTAAATTTATCAATGCAGATACC	926		
Db	601	AGTCTACATAGCATATAGCATTTCTACATTTGGGTTTAAATTTATCAATGCAGATACC	660		
Qy	927	TTTTGGGATTTCCAGCCCAATCAGACAGTGAACACATGGCAGCTGAGGTTCTTTC	986		
Db	661	TTTTGGGATTTCCAGCCCAATCAGACAGTGAACACATGGCAGCTGAGGTTCTTTC	720		
Qy	987	ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGACCGGATTAACAAACACAGAGTT	1046		
Db	721	ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGACCGGATTAACAAACACAGAGTT	780		
Qy	1047	CCAGACCTTTCTTTTGGGTTATCACTAGCTGAGGCTGCTGCTGCTGCTGCTGCTGCT	1106		
Db	781	CCAGACCTTTCTTTTGGGTTATCACTAGCTGAGGCTGCTGCTGCTGCTGCTGCTGCT	840		
Qy	1107	CTATTGACTTATACAGGTTTACATTCACCACTGAGTGGCAGGTTTATTCATTGTGGGA	1166		
Db	841	CTATTGACTTATACAGGTTTACATTCACCACTGAGTGGCAGGTTTATTCATTGTGGGA	900		
Qy	1167	TACTGGGATGAAAAATACACATTTCCAAATTTTGCATCAGTGTCTGAGCATCAACCTTAC	1226		
Db	901	TACTGGGATGAAAAATACACATTTCCAAATTTTGCATCAGTGTCTGAGCATCAACCTTAC	960		
Qy	1227	GACTTGGGTTCTTTCTTTCTTTGATCTACATATCTTGTATGATCTTCCAGAGGCT	1286		
Db	961	GACTTGGGTTCTTTCTTTCTTTGATCTACATATCTTGTATGATCTTCCAGAGGCT	1020		
Qy	1287	TTGGTTCTGCATCAAAAATATCAACGATGAAGA	1320		
Db	1021	TTGGTTCTGCATCAAAAATATCAACGATGAAGA	1054		

Db 2281 ACATCAGAACACTGGCTTGTAGGATATATAAAGTAAAGCAGCTGATTAACAGGAGACA 2340
 QY 2341 TTAGATCAAACTCGAGTCACCAACATTTCCCAAAACAGAGATTTTGTCAAAAGAG 2400
 Db 2341 TTAGATCAAACTCGAGTCACCAACATTTCCCAAAACAGAGATTTTGTCAAAAGAG 2400
 QY 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAAAGGCAAGAA 2460
 Db 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAAAGGCAAGAA 2460
 QY 2461 ATATCTAAGAGACTGTTTAA 2481
 Db 2461 ATATCTAAGAGACTGTTTAA 2481

RESULT 2
 US-09-945-527-62
 ; Sequence 62, Application US/09945527
 ; Publication No. US2003005588A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Robison, Keith E.
 ; TITLE OF INVENTION: Nucleic Acid and Protein Homologs
 ; FILE REFERENCE: 35900/23795
 ; CURRENT APPLICATION NUMBER: US/09/945,527
 ; PRIORITY FILING DATE: 2001-08-29
 ; NUMBER OF SEQ ID NOS: 65
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 62
 ; LENGTH: 1828
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: 1767, 1771
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-945-527-62

Query Match 43.1%; Score 1069; DB 10; Length 1828;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1069; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1413 GTCTGCAATGCTTTTCAATGTTTGGAGCACTATTGGGGGATGACATGAAAGGGA 1472
 Db 69 GTCTGCAATGCTTTTCAATGTTTGGAGCACTATTGGGGGATGACATGAAAGGGA 128
 QY 1473 AAATCCACTGTGGAGGACAGCAGTGTAGGATGACAAAGAAACCAAGGAAATTTGTA 1532
 Db 129 AAATCCACTGTGGAGGACAGCAGTGTAGGATGACAAAGAAACCAAGGAAATTTGTA 188
 QY 1533 TGATAGGAGGTTAAAGTCAGGAAACATCACTGAAAGGAAACCAAGGAGGAT 1592
 Db 189 TGATAGGAGGTTAAAGTCAGGAAACATCACTGAAAGGAAACCAAGGAGGAT 248
 QY 1593 AGSCCTTAATAAAGCAATGTCACCATGTTGATGCTGATGCTGATGATGATGCTG 1652
 Db 249 AGSCCTTAATAAAGCAATGTCACCATGTTGATGCTGATGCTGATGATGATGCTG 308
 QY 1653 TGTCCTACTGCTGGGTCACAGCAATGCTGATGCTGATGCTGATGCTGATGCTGCTC 1712
 Db 309 TGTCCTACTGCTGGGTCACAGCAATGCTGATGCTGATGCTGATGCTGATGCTGCTC 368
 QY 1713 ATACAATCATGATGGCACCAGGAATATCTTAGATGATTTAGAGAGCTTACTTTGGCT 1772
 Db 369 ATACAATCATGATGGCACCAGGAATATCTTAGATGATTTAGAGAGCTTACTTTGGCT 428
 QY 1773 AAGGCAAAATACAGATGACATGACAGATGATGCTGCTGGGATTTAGCTATCAGAT 1832
 Db 429 AAGGCAAAATACAGATGACATGACAGATGATGCTGCTGGGATTTAGCTATCAGAT 488
 QY 1833 AGCTGGAATGGCTTAATAGAACTACGTTGGTGGGATTAATACCTGGAATAACAGCCACAT 1892

Db 489 AGCTGGAATGGCTTAATAGAACTACGTTGGTGGATTAATAACACCTGGAATAACAGCCACAT 548
 QY 1893 AGCACTGGTGGGAAAGCTATGCTTCTTAATGAACACAGCAGCTATAAATCATAGGAC 1952
 Db 549 AGCACTGGTGGGAAAGCTATGCTTCTTAATGAACACAGCAGCTATAAATCATAGGAC 608
 QY 1953 TCTAGATGTAGATTAATGTTTGGTATTATTTTGGAGGGGTTATTTGGCTATTTCTGGTGATGA 2012
 Db 609 TCTAGATGTAGATTAATGTTTGGTATTATTTTGGAGGGGTTATTTGGCTATTTCTGGTGATGA 668
 QY 2013 TATCAACAAATTTCTCTGATGTTTAGGATGCTGAAAGGAGAACATCCCAAGACATTCG 2072
 Db 569 TATCAACAAATTTCTCTGATGTTTAGGATGCTGAAAGGAGAACATCCCAAGACATTCG 728
 QY 2073 GGAAGTGACTATTTTACCCCAAGGAGAAATTCGTTAGACAAAGCAGGATCCCTTAC 2132
 Db 729 GGAAGTGACTATTTTACCCCAAGGAGAAATTCGTTAGACAAAGCAGGATCCCTTAC 788
 QY 2133 TTTGTTGAATTTGCTTATGTATATAAATGTCATACATACAGATTTGGAGAAATGCAGCTGGA 2192
 Db 789 TTTGTTGAATTTGCTTATGTATATAAATGTCATACATACAGATTTGGAGAAATGCAGCTGGA 848
 QY 2193 TTTTCTCACACCCCAAGGTTTGAACGAAACACGTAATGCTGAGATTTGGAATTAAGACAT 2252
 Db 849 TTTTCTCACACCCCAAGGTTTGAACGAAACACGTAATGCTGAGATTTGGAATTAAGACAT 908
 QY 2253 TAAATTCACAACTTTGGAAGAGCCCTTTACATCAGAACACCTGGCTTTAGGATATATAA 2312
 Db 909 TAAATTCACAACTTTGGAAGAGCCCTTTACATCAGAACACCTGGCTTTAGGATATATAA 968
 QY 2313 AGTAAAGCAGCTGATTAACAGGAGACATTTAGATCACAACCTCGAGTCACCAACATTTT 2372
 Db 969 AGTAAAGCAGCTGATTAACAGGAGACATTTAGATCACAACCTCGAGTCACCAACATTTT 1028
 QY 2373 CCACAAACAGAGATTTTGTCAAAGAGACTACCAAGAGAGCGTGGCTACATTAATAA 2432
 Db 1029 CCACAAACAGAGATTTTGTCAAAGAGACTACCAAGAGAGCGTGGCTACATTAATAA 1088
 QY 2433 TAAGCTGGTTTTAAAGAAAGCAAGAAATATCTTAAGAGAGCTGTTTAA 2481
 Db 1089 TAAGCTGGTTTTAAAGAAAGCAAGAAATATCTTAAGAGAGCTGTTTAA 1137

RESULT 3
 US-10-106-698-330
 ; Sequence 330, Application US/10106698
 ; Publication No. US20030109690A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
 ; FILE REFERENCE: PA005P1
 ; CURRENT APPLICATION NUMBER: US/10/106,698
 ; PRIORITY FILING DATE: 2002-03-27
 ; PRIOR FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524
 ; PRIOR FILING DATE: 1999-09-29
 ; PRIOR APPLICATION NUMBER: US 60/157,137
 ; PRIOR FILING DATE: 1999-11-03
 ; PRIOR APPLICATION NUMBER: US 60/163,280
 ; NUMBER OF SEQ ID NOS: 8564
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 330
 ; LENGTH: 1209
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1203) (1203)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-10-106-698-330

Query Match 42.5%; Score 1054; DB 15; Length 1209;
 Best Local Similarity 100.0%; Pred. No. 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2004, 00:55:34 ; Search time 2195.32 Seconds
(without alignments)
6239.285 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagccctcgccccc.....tatctaagaagactgttaa 2481

Scoring table: OLIGO_NUC
Gapco 60.0 , Gapext 60.0

Searched: 4093002 seqs, 2760418825 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8186004

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 2: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 3: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 4: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 5: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 6: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 7: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 8: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 9: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 10: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 11: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 12: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 13: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 14: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 15: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 16: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 17: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 18: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 19: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 20: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:
- 21: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2481	100.0	2481	15	US-10-028-384-1
2	1089	43.1	1828	10	US-09-945-527-62
3	1054	42.5	1209	15	US-10-106-698-330
4	1054	42.5	1209	16	US-10-264-237-412
5	1054	42.5	1543	10	US-09-974-879-133
6	1054	42.5	1543	10	US-09-305-736-133
7	1054	42.5	1543	10	US-09-818-683-133
8	1054	42.5	1543	11	US-09-818-683-133
9	1054	42.5	1543	16	US-10-621-401-133
10	840	33.9	2660	16	US-10-264-049-630
11	653	26.3	1114	16	US-10-296-115-629
12	500	20.2	500	9	US-09-998-598-1643

13	430	17.3	430	9	US-09-878-178-717	Sequence 717, App
14	430	17.3	430	13	US-10-046-935-717	Sequence 717, App
15	430	17.3	430	14	US-10-146-502-717	Sequence 717, App
16	402	16.2	558	14	US-10-052-283-433	Sequence 433, App
17	359	14.5	387	16	US-10-276-774-173	Sequence 173, App
18	351	14.1	616	9	US-09-879-536-332	Sequence 332, App
19	328	13.2	485	10	US-09-918-995-11283	Sequence 11283, A
c	20	12.7	743	10	US-09-945-527-63	Sequence 63, Appl
21	179	7.2	474	14	US-10-052-283-454	Sequence 454, App
22	156	6.3	333	16	US-10-276-774-323	Sequence 323, App
c	23	5.4	312	9	US-09-796-692-8870	Sequence 8870, Ap
c	24	5.4	312	14	US-10-040-862-8870	Sequence 8870, Ap
c	25	5.4	312	16	US-10-057-475B-8870	Sequence 8870, Ap
c	26	5.4	312	16	US-10-154-844B-8870	Sequence 8870, Ap
c	27	5.4	312	17	US-10-764-324-8870	Sequence 8870, Ap
c	28	5.0	352	9	US-09-878-178-1032	Sequence 1032, Ap
c	29	5.0	352	13	US-10-046-935-1032	Sequence 1032, Ap
c	30	5.0	352	14	US-10-146-502-1032	Sequence 1032, Ap
c	31	4.2	483	10	US-09-918-995-23896	Sequence 23896, A
c	32	2.6	2710	15	US-10-028-384-3	Sequence 3, Appli
c	33	1.8	637	15	US-10-002-631C-183	Sequence 183, App
c	34	1.8	640	15	US-10-002-631C-179	Sequence 179, App
c	35	1.1	424	10	US-09-918-995-17769	Sequence 17769, A
c	36	1.0	235	10	US-09-945-527-61	Sequence 61, Appl
c	37	0.9	252	18	US-10-128-558-54	Sequence 54, Appl
c	38	0.9	307	9	US-09-736-457-375	Sequence 375, App
c	39	0.9	307	9	US-09-736-457-588	Sequence 588, App
c	40	0.9	307	9	US-09-736-457-1271	Sequence 1271, Ap
c	41	0.9	307	9	US-09-902-941-375	Sequence 375, App
c	42	0.9	307	9	US-09-902-941-588	Sequence 588, App
c	43	0.9	307	9	US-09-902-941-1271	Sequence 1271, Ap
c	44	0.9	307	9	US-09-849-626-375	Sequence 375, App
c	45	0.9	307	9	US-09-849-626-588	Sequence 588, App

ALIGNMENTS

RESULT 1
US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPTIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)
; OTHER INFORMATION:
US-10-028-384-1

Query Match	100.0%;	Score	2481;	DB	15;	Length	2481;
Best Local Similarity	100.0%;	Pred. No.	0;				
Mismatches	0;	Conservative	0;	Indels	0;	Gaps	0;
Qy	1	ATGCGGAGCCCTCGCCCGGAGCAAGCAAGTCGCTCCCTCAACTGCTCCCGTGG	60				
Db	1	ATGCGGAGCCCTCGCCCGGAGCAAGTCGCTCCCTCAACTGCTCCCGTGG	60				
Qy	61	AGTGCGCTCATGCGCCCTCGGAAACAGCGGCGCGGCGCCCGAGTGC	120				
Db	61	AGTGCGCTCATGCGCCCTCGGAAACAGCGGCGCGGCGCCCGAGTGC	120				

~~DUPLICATE~~

Blank page

```
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 588
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-588
```

```
Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
Db 129 TGGTGGGATTATGGCTATCAGAT 151
```

```
RESULT 13
US-09-614-124B-1271
; Sequence 1271, Application US/09614124B
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1271
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-1271
```

```
Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
Db 129 TGGTGGGATTATGGCTATCAGAT 151
```

```
RESULT 14
US-09-671-325-375
; Sequence 375, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
```

```
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 375
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-671-325-375
```

```
Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
Db 129 TGGTGGGATTATGGCTATCAGAT 151
```

```
RESULT 15
US-09-671-325-588
; Sequence 588, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 588
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-671-325-588
```

```
Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
Db 129 TGGTGGGATTATGGCTATCAGAT 151
```

Search completed: December 16, 2004, 07:21:29
Job time : 260.852 secs

; APPLICANT: Lodes, Michael A.
 ; APPLICANT: Fanger, Gary
 ; APPLICANT: Vedvick, Tom
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Fan, Liqun
 ; APPLICANT: Wang, Aijun
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.478C15
 ; CURRENT APPLICATION NUMBER: US/09/736,457
 ; CURRENT FILING DATE: 2000-12-13
 ; NUMBER OF SEQ ID NOS: 1864
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 375
 ; LENGTH: 307
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-09-736-457-375

Query Match 0.9%; Score 23; DB 4; Length 307;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
 DB 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 9
 US-09-736-457-588
 ; Sequence 588, Application US/09736457
 ; Patent No. 6509448
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Lodes, Michael A.
 ; APPLICANT: Fanger, Gary
 ; APPLICANT: Vedvick, Tom
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Fan, Liqun
 ; APPLICANT: Wang, Aijun
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.478C15
 ; CURRENT APPLICATION NUMBER: US/09/736,457
 ; CURRENT FILING DATE: 2000-12-13
 ; NUMBER OF SEQ ID NOS: 1864
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 588
 ; LENGTH: 307
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-09-736-457-588

Query Match 0.9%; Score 23; DB 4; Length 307;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
 DB 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 10
 US-09-736-457-1271
 ; Sequence 1271, Application US/09736457
 ; Patent No. 6509448
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Lodes, Michael A.
 ; APPLICANT: Fanger, Gary
 ; APPLICANT: Vedvick, Tom
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Fan, Liqun
 ; APPLICANT: Wang, Aijun
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.478C15
 ; CURRENT APPLICATION NUMBER: US/09/736,457
 ; CURRENT FILING DATE: 2000-12-13
 ; NUMBER OF SEQ ID NOS: 1864
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1271
 ; LENGTH: 307
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-09-736-457-1271

Query Match 0.9%; Score 23; DB 4; Length 307;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
 DB 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 11
 US-09-614-124B-375
 ; Sequence 375, Application US/09614124B
 ; Patent No. 6630574
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Lodes, Michael A.
 ; APPLICANT: Fanger, Gary
 ; APPLICANT: Vedvick, Tom
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Mannion, Jane
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
 ; FILE REFERENCE: 210121.478C9
 ; CURRENT APPLICATION NUMBER: US/09/614,124B
 ; CURRENT FILING DATE: 2001-07-11
 ; NUMBER OF SEQ ID NOS: 1668
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 375
 ; LENGTH: 307
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-09-614-124B-375

Query Match 0.9%; Score 23; DB 4; Length 307;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
 DB 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 12
 US-09-614-124B-588
 ; Sequence 588, Application US/09614124B
 ; Patent No. 6630574
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Bangur, Chaitanya S.

;; PRIOR APPLICATION NUMBER: US 60/122,487
;; PRIOR FILING DATE: 1999-02-26
;; NUMBER OF SEQ ID NOS: 36681
;; SOFTWARE: Patent.pm
;; SEQ ID NO 21090
;; LENGTH: 245
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-513-999C-21090

Query Match 9.5%; Score 236; DB 4; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.2e-112;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 CTACTTTCCCTGCTTACAAGAGAACTTTGGAACCAAGAGCAGGACTTTAGCTGCTGTT 610
DB 1 CTACTTTCCCTGCTTACAAGAGAACTTTGGAACCAAGAGCAGGACTTTAGCTGCTGTT 60

QY 611 TTATTGCTATTGTACAGGCTACATATCTCGGTGCTAGTCTGGATCCCTTTGATAATGAAG 670
DB 61 TTATTGCTATTGTACAGGCTACATATCTCGGTGCTAGTCTGGATCCCTTTGATAATGAAG 120

QY 671 GCATTGCTATTGTGCACTTCAGTTCACATACACTATTATGGGTAAAACTCTGTAAGAACTG 730
DB 121 GCATTGCTATTGTGCACTTCAGTTCACATACACTATTATGGGTAAAACTCTGTAAGAACTG 180

QY 731 GGTCAAGTTTTTGGACAAATGTGCTGCTTATCTTATCTTATATGCTCTCTGCT 786
DB 181 GGTCAAGTTTTTGGACAAATGTGCTGCTTATCTTATCTTATATGCTCTCTGCT 236

RESULT 5

US-09-702-705-375
;; Sequence 375, Application US/09702705
;; Patent No. 6504010
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Tongtong
;; APPLICANT: Bangur, Chaitanya S.
;; APPLICANT: Lodes, Michael A.
;; APPLICANT: Fanger, Gary
;; APPLICANT: Vedvick, Tom
;; APPLICANT: Carter, Darrick
;; APPLICANT: Retter, Marc
;; APPLICANT: Mannion, Jane
;; APPLICANT: Fan, Liqun
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
;; FILE REFERENCE: 210121.478C14
;; CURRENT APPLICATION NUMBER: US/09/702,705
;; CURRENT FILING DATE: 2000-10-30
;; NUMBER OF SEQ ID NOS: 1833
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 375
;; LENGTH: 307
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-09-702-705-375

Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
DB 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 6

US-09-702-705-588
;; Sequence 588, Application US/09702705
;; Patent No. 6504010
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Tongtong

;; APPLICANT: Bangur, Chaitanya S.
;; APPLICANT: Lodes, Michael A.
;; APPLICANT: Fanger, Gary
;; APPLICANT: Vedvick, Tom
;; APPLICANT: Carter, Darrick
;; APPLICANT: Retter, Marc
;; APPLICANT: Mannion, Jane
;; APPLICANT: Fan, Liqun
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
;; FILE REFERENCE: 210121.478C14
;; CURRENT APPLICATION NUMBER: US/09/702,705
;; CURRENT FILING DATE: 2000-10-30
;; NUMBER OF SEQ ID NOS: 1833
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 588
;; LENGTH: 307
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-09-702-705-588

Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
DB 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 7

US-09-702-705-1271
;; Sequence 1271, Application US/09702705
;; Patent No. 6504010
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Tongtong
;; APPLICANT: Bangur, Chaitanya S.
;; APPLICANT: Lodes, Michael A.
;; APPLICANT: Fanger, Gary
;; APPLICANT: Vedvick, Tom
;; APPLICANT: Carter, Darrick
;; APPLICANT: Retter, Marc
;; APPLICANT: Mannion, Jane
;; APPLICANT: Fan, Liqun
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
;; FILE REFERENCE: 210121.478C14
;; CURRENT APPLICATION NUMBER: US/09/702,705
;; CURRENT FILING DATE: 2000-10-30
;; NUMBER OF SEQ ID NOS: 1833
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 1271
;; LENGTH: 307
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-09-702-705-1271

Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
DB 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 8

US-09-736-457-375
;; Sequence 375, Application US/09736457
;; Patent No. 6509448
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Tongtong
;; APPLICANT: Bangur, Chaitanya S.

QY	1983	TGAGGGGTTATTGGCTATTCTGGTGATGATCAACAAATTTCTCTGGATGTTAGGAT	2042
Db	301	TGAGGGGTTATTGGCTATTCTGGTGATGATCAACAAATTTCTCTGGATGTTAGGAT	360
QY	2043	AGCTGAAGGAGACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCACAGGGAGA	2102
Db	361	AGCTGAAGGAGAAATCCCAAGACATTCGGGAAAGTGACTATTTTACCCACAGGGAGA	420
QY	2103	ATTCCGTGTAGACAAAGCAGGATCCCTACTTTGTTGAAATGCGCTTATGATATAAATGTC	2162
Db	421	ATTCCGTGTAGACAAAGCAGGATCCCTACTTTGTTGAAATGCGCTTATGATATAAATGTC	480
QY	2163	ATACTACAGATTTGGAGAAATGCAGCT	2189
Db	481	ATACTACAGATTTGGAGAAATGCAGCT	507

RESIST. 2

```

US-09-328-111-332
; Sequence 332, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Asle, Jon H.
; APPLICANT: Burgees, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328.111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 332
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(616)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-332

```

Query Match	14.1%;	Score 351;	DB 3;	Length 616;
Best Local Similarity	100.0%;	Pred. No. 1.5e-172;		

622	GTACAGGCTACATATCTCGGTGAGTACGTCGGATCCTTTGATTAATCAAGGCGATTGCTATT	681
	5 GTACAGGCTACATATCTCGGTGAGTACGTCGGATCCTTTGATTAATCAAGGCGATTGCTATT	64
682	TTTTCGACTTCAGTTTCACATACCTATTATGGGTAAAACTGTAAAAACTCGGTGCGATTTTT	741
	65 TTTTCGACTTCAGTTTCACATACCTATTATGGGTAAAACTGTAAAAACTCGGTGCGATTTTT	124
742	TGGACAATGTGCTGCTCTTATCCTATTCTATATGGTCTCTGCTTGGGTGGTATTATGTA	801
	125 TGGACAATGTGCTGCTCTTATCCTATTCTATATGGTCTCTGCTTGGGTGGTATTATGTA	184
802	TTTATCATCAATCTTATTCGACTGCATGTAATTTGTGTGTTACTGATCAGAGATACAGC	861
	185 TTTATCATCAATCTTATTCGACTGCATGTAATTTGTGTGTTACTGATCAGAGATACAGC	244

862	AAAGAGTCTCATAGCATATAGCACTTTCACATTGTGGTTTAAATATTATCAATGCAG	921
Cy		
245	AAAGAGTCTCATAGCATATAGCACTTTCACATTGTGGTTTAAATATTATCAATGCAG	304
Db		
922	ATACCTTTTGTGGGATTCACGCCAATCAGAACCAAGTGAACACATGCCAGCT	972
Cy		
305	ATACCTTTTGTGGGATTCACGCCAATCAGAACCAAGTGAACACATGCCAGCT	355
Db		

RESULT 3

```

US-09-513-999C-1438
; Sequence 1438, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Wilbe Edwards, J.B.
; APPLICANT: Duclert, A. Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1438
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 54..248
US-09-513-999C-1438

```

Query Match	10.1%;	Score 250;	DB 4;	Length 250;
Best Local Similarity	100.0%;	Pred. No. 6.2e-120;		

794	QY	GTTATGTTATTATCATCAATCTTATTCCACTGCATGTTTGTGTTGTTACTGATGCAGA	853
	Dd		
1		GTTATGTTATTATCATCAATCTTATTCCACTGCATGTTTGTGTTGTTACTGATGCAGA	60
854	QY	GATCAGCMAAAGAGTCTCATAGCATATAGACATTCTACATTTGGGTTTAAATTAT	913
	Dd		
61		GATCAGCMAAAGAGTCTCATAGCATATAGACATTCTACATTTGGGTTTAAATTAT	120
914	QY	CAATCGAGATACCTTTTGTGGGATTCCAGCCAACTCAGAACACATGGCAGCTG	973
	Dd		
121		CAATCGAGATACCTTTTGTGGGATTCCAGCCAACTCAGAACACATGGCAGCTG	180
974	QY	CAGTGTCTTTTGATTTGCTCCAGCTTATGCTTTCTTGCAGTATCTGAGAGACCGATTAA	1033
	Dd		
181		CAGTGTCTTTTGATTTGCTCCAGCTTATGCTTTCTTGCAGTATCTGAGAGACCGATTAA	240
1034	QY	CAAAACAAGA	1043
241	Dd	CAAAACAAGA	250

DEPT. T. A

```

RESULT 4
US-09-513-9990
; Sequence 21090, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A. J.Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 10:34:04 ; Search time 259.852 Seconds
(without alignments)
6786.440 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagccctggcccc.....tatctaagaagactgtttaa 2481

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	507	20.4	507	US-09-513-999C-1965	Sequence 1965, Ap
2	351	14.1	616	US-09-328-111-332	Sequence 332, App
3	250	10.1	250	US-09-513-999C-1438	Sequence 1438, Ap
4	236	9.5	245	US-09-513-999C-21090	Sequence 21090, A
5	23	0.9	307	US-09-702-705-375	Sequence 375, App
6	23	0.9	307	US-09-702-705-588	Sequence 588, App
7	23	0.9	307	US-09-702-705-1271	Sequence 1271, Ap
8	23	0.9	307	US-09-736-457-375	Sequence 375, App
9	23	0.9	307	US-09-736-457-588	Sequence 588, App
10	23	0.9	307	US-09-736-457-1271	Sequence 1271, Ap
11	23	0.9	307	US-09-614-124B-375	Sequence 375, App
12	23	0.9	307	US-09-614-124B-588	Sequence 588, App
13	23	0.9	307	US-09-614-124B-1271	Sequence 1271, Ap
14	23	0.9	307	US-09-671-325-375	Sequence 375, App
15	23	0.9	307	US-09-671-325-588	Sequence 588, App
16	23	0.9	307	US-09-671-325-1271	Sequence 1271, Ap
17	23	0.9	307	US-09-589-184-375	Sequence 375, App
18	23	0.9	307	US-09-589-184-588	Sequence 588, App
19	23	0.9	307	US-09-589-184-1271	Sequence 1271, Ap
20	23	0.9	307	US-09-658-824-375	Sequence 375, App
21	23	0.9	307	US-09-658-824-588	Sequence 588, App
22	23	0.9	307	US-09-658-824-1271	Sequence 1271, Ap
23	23	0.9	487	US-09-702-705-1655	Sequence 1655, Ap
24	23	0.9	487	US-09-736-457-1655	Sequence 1655, Ap
25	23	0.9	487	US-09-614-124B-1655	Sequence 1655, Ap
26	23	0.9	487	US-09-671-325-1655	Sequence 1655, Ap
27	21	0.8	480	US-09-489-039A-5999	Sequence 5999, Ap

ALIGNMENTS

RESULT 1

US-09-513-999C-1965
Sequence 1965, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
PATENT NO. 6783961
FILE REFERENCE: 59,US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 1965
LENGTH: 507
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 122..505
US-09-513-999C-1965

Query Match 20.4%; Score 507; DB 4; Length 507;
Best Local Similarity 100.0%; Pred. No. 7,1e-254;
Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1683	CTACTCTAGTCCAAAGTGTAGTCTGGCTCATCAATCATGATGGCCACAGGATATCTT	1742
Db	1	CTACTCTAGTCCAAAGTGTAGTCTGGCTCATCAATCATGATGGCCACAGGATATCTT	60
Qy	1743	AGATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAGAGT	1802
Db	61	AGATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAGAGT	120
Qy	1803	AATGCTCTGGTGGGATTTATGGCTATCAGATAGTGGGATGGCTTAATAGAACTACGTTGGT	1862
Db	121	AATGCTCTGGTGGGATTTATGGCTATCAGATAGTGGGATGGCTTAATAGAACTACGTTGGT	180
Qy	1863	GGATAATAACACCTGGGATTAACAGCCACATAGACACTGGTGGGAAAAGCTATGCTCTTAA	1922
Db	181	GGATAATAACACCTGGGATTAACAGCCACATAGACACTGGTGGGAAAAGCTATGCTCTTAA	240
Qy	1923	TGAACACAGCCCTTATAAATCATAGGACTCTAGATGTAGATTTGTTTGGTTATTTT	1982
Db	241	TGAACACAGCCCTTATAAATCATAGGACTCTAGATGTAGATTTGTTTGGTTATTTT	300

Blank page
~~duplicate~~

581	Db	TTTGTAATTCGCTTATGTTATAAAATGTCATACTACAGATTGGAGAAATGCGACTGGAT	640
2194	Qy	TTTCGTACACCCCCAGGCTTTTGACCGAACACAGTAATGCTGAGATTGGAATTAAGGACATT	2253
641	Db	TTTCGTACACCCCCAGGCTTTTGACCGAACACAGTAATGCTGAGATTGGAATTAAGGACATT	700
2254	Qy	AAATTTCAACATTTTGGAAGAAGCGCTTTTACATCAGAACTGCGCTTGTTTAGGATATATAA	2313
701	Db	AAATTTCAACATTTTGGAAGAAGCGCTTTTACATCAGAACTGCGCTTGTTTAGGATATATAA	760
2314	Qy	GTAAGACGCTGNTAACGGGAGACATTAGATATCBAACCTCGAGTCACCACTTTTC	2373
761	Db	GTAAGACGCTGNTAACGGGAGACATTAGATATCBAACCTCGAGTCACCACTTTTC	820
2374	Qy	CCAAAACAGAGTATTTGTCCTAAAGAGAGCTACAAAAGGAAGCGTGGCTACATTAAAAAT	2433
821	Db	CCAAAACAGAGTATTTGTCCTAAAGAGAGCTACAAAAGGAAGCGTGGCTACATTAAAAAT	880
2434	Qy	AAGCTGTTTTTTAAGAAGCAAGAAAAATATCTAAGAAGACTGTTTTAA	2481
881	Db	AAGCTGTTTTTTAAGAAGCAAGAAAAATATCTAAGAAGACTGTTTTAA	928

Search completed: December 16, 2004, 00:55:22
Job time : 16566.4 secs

VERSION	BD137303.1	GI:223232248
KEYWORDS	JP 2002505878-A/64.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. Specht, T., Hinzmann, B., Schmitt, A., Pilarsky, C., Dahl, E. and Rosenthal, A. (Bases 1 to 2503)	
AUTHORS	Human nucleic acid sequence originating in prostatic tissue Patent: JP 200305878-A 64 26-FEB-2002; METAGEN GESELLSCHAFT FUER GENOM FORSCHUNG MBH CS Homo sapiens (human) PN JP 2002505878-A/64 PD 26-FEB-2002 PF 09-MAR-1999 JP 2000535742 PI 10-MAR-1998 DE 198 11 194.0 PR THOMAS SPECHT, BERND HINZMANN, ARMIN SCHMITT, CHRISTIAN PILARSKY, PI EDGAR DAHL, PI ANDRE ROSENTHAL PC C12N15/09,A61K38/00,A61K39/395,A61K39/395,A61K48/00,A61P13/08, PC A61P35/00,C12N15/00,C12N11/19,C12N1/21,C12N5/10,C12N15/02, PC PC C07K14/47,C07K16/18,C12N1/19,C12N1/21,C12N5/10,C12N15/02, PC C12P21/02,	
TITLE	JOURNAL	
COMMENT	Key Human nucleic acid sequence originating in prostatic tissue PH Location/Qualifiers FT source 1..2503 FT /organism='Homo sapiens (human)'. FEATURES source 1..2503 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"	
ORIGIN	Query Match 33.4%; Score 828; DB 6; Length 2503; Best Local Similarity 100.0%; Pred. No. 0; Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1654 GTCCACTGTACCTGGGTCACAAGCAATGCCTACTCTAGTCCAAGTGTAAGTCCTGGCTCA 1713	
DB	101 GTCCACTGTACCTGGGTCACAAGCAATGCCTACTCTAGTCCAAGTGTAAGTCCTGGCTCA 160	
QY	1714 TACAATCATGNTGCCACCAGGAATCTTAGATGATTTAGAGAAGCTTACTTTTGGCTA 1773	
DB	161 TACAATCATGNTGCCACCAGGAATCTTAGATGATTTAGAGAAGCTTACTTTTGGCTA 220	
QY	1774 AGGCAAAATCAGATGAACATGCACGAGTAATGTCTTGTGGGATTTATGGCTATCAGATA 1893	
DB	221 AGGCAAAATCAGATGAACATGCACGAGTAATGTCTTGTGGGATTTATGGCTATCAGATA 280	
QY	1834 CTGTGAATGGCTAATGAACCTACGTTGGTGGATATAACACCTCGAAATAACAGCCACATA 1893	
DB	281 CTGTGAATGGCTAATGAACCTACGTTGGTGGATATAACACCTCGAAATAACAGCCACATA 340	
QY	1894 GCACGTGGTGGAAAAGCTATGTCTTCTAATGAAAACAGCGCCTATAAAATCATGAGACT 1953	
DB	341 GCACGTGGTGGAAAAGCTATGTCTTCTAATGAAAACAGCGCCTATAAAATCATGAGACT 400	
QY	1954 CTAGATGTAGATTATGTTTTTGGTTATTTTGGAGGGGTTATTGGCTATTCTTGGTGTGAT 2013	
DB	401 CTAGATGTAGATTATGTTTTTGGTTATTTTGGAGGGGTTATTGGCTATTCTTGGTGTGAT 460	
QY	2014 ATCACAAATTTCTCTGGATGGTTAGATAGCTGTAAGGAGACATCCCARAAGACATTCGG 2073	
DB	461 ATCACAAATTTCTCTGGATGGTTAGATAGCTGTAAGGAGACATCCCARAAGACATTCGG 520	
QY	2074 GAAGTGACTATTTTACCCACAGGGAGAAATTCGGTGTAGCAAAGCAGAGATCCCTACT 2133	
DB	521 GAAAGTGACTATTTTACCCACAGGGAGAAATTCGGTGTAGCAAAGCAGAGATCCCTACT 580	
QY	2134 TTGTTGAATTCGCTTATGTATAAAATGTCACTACTACAGATTTGGAGAAATCAGCTGGAT 2193	

[illegible]


```

Db      663 ACATCAGAACACTGGCTGTTAGGATATATAAAGTAAAGACACCTGATTAACAGGGAGACA 722
Qy      2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATATTGTCAAGAG 2400
Db      723 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATATTGTCAAGAG 782
Qy      2401 ACTACCAAAAGGAGCGTGGCTACATTAAATAAAGCTGTTTAAAGAAAGGCAAGAAA 2460
Db      783 ACTACCAAAAGGAGCGTGGCTACATTAAATAAAGCTGTTTAAAGAAAGGCAAGAAA 842
Qy      2461 ATATCTAAGAAAGCTGTTTAA 2481
Db      843 ATATCTAAGAAAGCTGTTTAA 863

RESULT 12
BD160013
LOCUS      2284 bp      DNA      linear      PAT 17-JAN-2003
DEFINITION      Primer for synthesizing full-length cDNA and use thereof.
ACCESSION      BD160013
VERSION      BD160013.1 GI:27865771
KEYWORDS      JP 2002191363-A/14856.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2284)
Ota,T., Isegai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 14856 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS      Homo sapiens (human)
PN      JP 2002191363-A/14856
PD      09-JUL-2002
PF      28-JUL-2000 JP 2002280990
PI      TOSHIO OTA,TAKAO ISOgai, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
        SAITO,
        JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
        KEIICHI NAGAI, TETSUJI OTSUKI
PC      C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
        10,
        C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT      CDS      Location/Qualifiers
SOURCE      1..2284
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match      34.7%; Score 861; DB 6; Length 2284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 861; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1621 ATGTTGATGCTGATGCTATTGATGATCTTTGCTGTCACGTACCTGGTCAAGCAAT 1680
Db      3 ATGTTGATGCTGATGCTATTGATGATCTTTGCTGTCACGTACCTGGTCAAGCAAT 62
Qy      1681 GCCTACTCTAGTCCAAAGTGTAGTCTCGCTCCTCATAATCATGATGGCCACCAAGGAATATC 1740
Db      63 GCCTACTCTAGTCCAAAGTGTAGTCTCGCTCCTCATAATCATGATGGCCACCAAGGAATATC 122
Qy      1741 TTAGATGATTTTACAGAGCTTACTTTTGGCTAAGCCAAATACAGATCAATCAGCA 1800
Db      123 TTAGATGATTTTACAGAGCTTACTTTTGGCTAAGCCAAATACAGATCAATCAGCA 182
Qy      1801 GTAATGCTTCGGTGGGATTAATGCTATCAGATAGCTGGAATGCTAATAGAACTACGTTG 1860
Db      183 GTAATGCTTCGGTGGGATTAATGCTATCAGATAGCTGGAATGCTAATAGAACTACGTTG 242

```

```

Qy      1861 GTGGATATATAACACCTCGAATAACAGCCACATAGCACTGGTGGAAAAAGCTATGCTTCT 1920
Db      243 GTGGATATATAACACCTCGAATAACAGCCACATAGCACTGGTGGAAAAAGCTATGCTTCT 302
Qy      1921 AATGAAAACAGAGCCTATAAATCATAGAGGACTCTAGATGATGATTTGTTGGTTAT 1980
Db      303 AATGAAAACAGAGCCTATAAATCATAGAGGACTCTAGATGATGATTTGTTGGTTAT 362
Qy      1981 TTTGGAGGGGTTATTGGCTATTCTGGTGATGATCAACAAATTTCTCTGGATGGTTAGG 2040
Db      363 TTTGGAGGGGTTATTGGCTATTCTGGTGATGATCAACAAATTTCTCTGGATGGTTAGG 422
Qy      2041 ATAGCTGAAGGAGAAATCCCAAGACATTCGGAAAGTCACTATTTCACCCACAGGGA 2100
Db      423 ATAGCTGAAGGAGAAATCCCAAGACATTCGGAAAGTCACTATTTCACCCACAGGGA 482
Qy      2101 GAATTCCTGGTGTAGACAAAGCAGGATCCCTACTTTTGTGTAATTGCTTATGTATAAAATG 2160
Db      483 GAATTCCTGGTGTAGACAAAGCAGGATCCCTACTTTTGTGTAATTGCTTATGTATAAAATG 542
Qy      2161 TCATACTACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCACAGGTTTGGACCGA 2220
Db      543 TCATACTACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCACAGGTTTGGACCGA 602
Qy      2221 ACACGTAATGCTGAGATTTGGAATTAAGACATTTAAATTCAAACATTTGGAAGAGGCTTT 2280
Db      603 ACACGTAATGCTGAGATTTGGAATTAAGACATTTAAATTCAAACATTTGGAAGAGGCTTT 662
Qy      2281 ACATCAGAACCTGCTGTTTGGATATATAAAGCAACCTGATACAGGGAGACA 2340
Db      663 ACATCAGAACCTGCTGTTTGGATATATAAAGCAACCTGATACAGGGAGACA 722
Qy      2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGTATTGTCAAGAAG 2400
Db      723 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGTATTGTCAAGAAG 782
Qy      2401 ACTACCAAAAGGAGCGTGGCTACATTAAATAAAGCTGTTTAAAGAAAGCAAGAAA 2460
Db      783 ACTACCAAAAGGAGCGTGGCTACATTAAATAAAGCTGTTTAAAGAAAGCAAGAAA 842
Qy      2461 ATATCTAAGAAAGCTGTTTAA 2481
Db      843 ATATCTAAGAAAGCTGTTTAA 863

RESULT 13
AK027789
LOCUS      2284 bp      mRNA      linear      PRI 30-JAN-2004
DEFINITION      Homo sapiens cDNA FLJ14883 fis, clone PLACE1003596, moderately
            similar to OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT.
ACCESSION      AK027789
VERSION      AK027789.1 GI:14042727
KEYWORDS      oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
Shitatori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
Kikkawa,E., Onura,Y., Abe,K., Kimihara,K., Katsuma,N., Sato,K.,
Tanikawa,M., Yamazaki,M., Nimomiyama,K., Ishibashi,T., Yamashita,H.,
Murakawa,M., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
Hirooka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
Yoshida,M., Hotsuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,

```

1588	GGATTAGGCCCTTAATATAAAAGCAATTGTCACCATGTTGTGATGCTGCTGATCTGCTATTGATGATG	1647
Qy		
61	GGATTAGGCCCTTAATATAAAAGCAATTGTCACCATGTTGTGATGCTGCTGATCTATTGATGATG	120
Ddb		
1648	TTTGCTCTCCACTGTACTCTGGGTCAACAAGCAATGCTACTCTAGTCCAAAGTGTAGTCCTG	1707
Qy		
121	TTTGCTCTGTCACCTGTACTCTGGGTCAACAAGCAATGCTACTCTAGTCCAAAGTGTAGTCCTG	180
Ddb		
1708	GCCTCATACAAATCATGATGGCACCAGGAATATCTTAGATGATTTTAGAAGCTTACTTTT	1767
Qy		
181	GCCTCATACAAATCATGATGGCACCAGGAATATCTTAGATGATTTTAGAAGCTTACTTTT	240
Ddb		
1768	TGGCTAAGGCAAAATACAGATGAACATGCACGAGTAATGTCTTGCTGGGATTATGGCTAT	1827
Qy		
241	TGGCTAAGGCAAAATACAGATGAACATGCACGAGTAATGTCTTGCTGGGATTATGGCTAT	300
Ddb		
1828	CAGATAGCTGGATGCTTAATGACTAGTACGTGGTGGATATAACACCTGGATATACAGC	1887
Qy		
301	CAGATAGCTGGATGCTTAATGACTAGTACGTGGTGGATATAACACCTGGATATACAGC	360
Ddb		
1888	CACATAGCACTGGTGGGAAAAGCTATGTTCTTAAATGAAAACAGCAGCCTATAAAATCATG	1947
Qy		
361	CACATAGCACTGGTGGGAAAAGCTATGTTCTTAAATGAAAACAGCAGCCTATAAAATCATG	420
Ddb		
1948	AGSACTCTAGATGTAGATTATGTTTGGTATTATTTTGGAGGGTTATTGGCTATTCCTGGT	2007
Qy		
421	AGSACTCTAGATGTAGATTATGTTTGGTATTATTTTGGAGGGTTATTGGCTATTCCTGGT	480
Ddb		
2008	GATGATATCAACAAATTTCTGATGTTTAGATAGCTGAAGGAGAAACATCCCAAGAC	2067
Qy		
481	GATGATATCAACAAATTTCTGATGTTTAGATAGCTGAAGGAGAAACATCCCAAGAC	540
Ddb		
2068	ATTCCGGAAAAGTGACTATTTTTACCACAGGAGAGAAATCCCGTGTAGACAAAGCAGAGATCC	2127
Qy		
541	ATTCCGGAAAAGTGACTATTTTTACCACAGGAGAGAAATCCCGTGTAGACAAAGCAGAGATCC	600
Ddb		
2128	CCTACTTTGTTGTAATGCTCTTATGTAATAAATGTATATACATACAGATTTGGAGAAATCGAG	2187
Qy		
601	CCTACTTTGTTGTAATGCTCTTATGTAATAAATGTATATACATACAGATTTGGAGAAATCGAG	660
Ddb		
2188	CTGATATTTTCGTACACCCCGAGTTTGTACCGAACACACGTAAATGCTGAGATTTGAAATTAAG	2247
Qy		
661	CTGATATTTTCGTACACCCCGAGTTTGTACCGAACACACGTAAATGCTGAGATTTGAAATTAAG	720
Ddb		
2248	GACATTAATAATCCAAAATTTGGAGAGAGCCTTTATCAGAAACACTGGCTTTGTAGGATA	2307
Qy		
721	GACATTAATAATCCAAAATTTGGAGAGAGCCTTTATCAGAAACACTGGCTTTGTAGGATA	780
Ddb		
2308	TATAAGTATAAAGCACACTGATTAACAGGGAGACATTTAGATCAACAACCTCGAGTCACCAAC	2367
Qy		
781	TATAAGTATAAAGCACACTGATTAACAGGGAGACATTTAGATCAACAACCTCGAGTCACCAAC	840
Ddb		
2368	ATTTTCCCAAAAACAGAAGTATTTGTCAAAAGAGACTACCAAAAGGAGCCTGGCTACATT	2427
Qy		
841	ATTTTCCCAAAAACAGAAGTATTTGTCAAAAGAGACTACCAAAAGGAGCCTGGCTACATT	900
Ddb		
2428	AAAAATAGCTGTTTTTTAAGAAAGCAAGAAAATATCTTAAGAAGACTGTTTAA	2481
Qy		
901	AAAAATAGCTGTTTTTTAAGAAAGCAAGAAAATATCTTAAGAAGACTGTTTAA	954
Ddb		

RESULT 11
AX8682932 LOCUS DNA
AX882932 Sequence 17837 from Patent EP1074517.
AX882932 AX882932
AX882932.1 GI:40037833
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1

Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
 Ishii,S., Sugiyama,T., Wakamatsu,A., Negai,K. and Otsuki,T.
 Primers for synthesising full-length cDNA and their use
 Patent: EP 1074617-A 17837 07-FEB-2001;
 Research Association for Biotechnology (JP)
 Location/Qualifiers
 1. .2284
 source
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 3. .863
 /notes="unnamed protein product"
 /codon_start=1
 /protein_id="CAE91783.1"
 /db_xref="GI:40037834"
 /translation="MLMLMLMPFAVHCTVWTSNAYSSPVVLASVNHSDGTRNILDFF
 RAYAFWIKMTDDEHAFVMSWDYDGYQAGMANETLTDVNTWNNSHIALVGVKAMSGNE
 TAAAYLKQNTLOVDVYLVFGVIGYDGDINDKFLMVRVIAEGHPKDIRESDFYFPQG
 EPRVDKAGSPITNCLMAYKMSYFREGMQLDFTPGFQRTNRNABIGNKDIKFKHLEE
 APTSDHVLARIYKVKAPDNRETLDPKRVNTIIPKQKYLSTKTKRKRCYIKNKLVPFK
 KKKKISKXTV"
 ORIGIN
 Query Match 34.7%; Score 861; DB 6; Length 2284;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 861; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1621 ATGTGATGCTGATGCTATTGATGATGTTTGCCTCCACTGTACCTGGGTCAACGAAT 1680
 3 ATGTGATGCTGATGCTATTGATGATGTTTGCCTCCACTGTACCTGGGTCAACGAAT 62
 1681 GCCTACTCTAGTCCBAGTGTTAGTCTGGCTCATACATCATGATGCGCACCGATATCT 1740
 63 GCCTACTCTAGTCCBAGTGTTAGTCTGGCTCATACATCATGATGCGCACCGATATCT 122
 1741 TTAGATGATTTAGAGAAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACGA 1800
 123 TTAGATGATTTAGAGAAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACGA 182
 1801 GTAAATGCTTTGGTGGGATTTAGCTATCAGATAGCTGAATGGCTAATAGAACTACGTTG 1860
 183 GTAAATGCTTTGGTGGGATTTAGCTATCAGATAGCTGAATGGCTAATAGAACTACGTTG 242
 1861 GTGGATTAATAACCTCGNATACAGCACATAGCACTGTTGGGAAAGCTATGCTCTCT 1920
 243 GTGGATTAATAACCTCGNATTAACAGCACATAGCACTGTTGGGAAAGCTATGCTCTCT 302
 1921 AATGAACACAGCAGCCTATAAAATCATAGAGCACTCTAGATGTAGATTATGTTTGGTTAT 1980
 303 AATGAACACAGCAGCCTATAAAATCATAGAGCACTCTAGATGTAGATTATGTTTGGTTAT 362
 1981 TTTGGAGGGGTTATTTGGCTATTTCGTGATCATATCAACAAATTTCTCGATGGTTAGG 2040
 363 TTTGGAGGGGTTATTTGGCTATTTCGTGATCATATCAACAAATTTCTCGATGGTTAGG 422
 2041 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGACATATTTTACCACACAGGA 2100
 423 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGACATATTTTACCACACAGGA 482
 2101 GAATTCCTGTGAGACAAAGCAGGATCCCTACTTTTGTGAATTCGCTTATGTATAAAATG 2160
 483 GAATTCCTGTGAGACAAAGCAGGATCCCTACTTTTGTGAATTCGCTTATGTATAAAATG 542
 2161 TCATATCTACAGATTTTGGAAATGACGTGGATTTTGTACACCCACAGGTTTGTACCGA 2220
 543 TCATATCTACAGATTTTGGAAATGACGTGGATTTTGTACACCCACAGGTTTGTACCGA 602
 2221 ACACGTAATGCTGAGATTTGGAAATTAAGACATTAATTAATTCAAACATTTGGAAGAGCCTTT 2280
 603 ACACGTAATGCTGAGATTTGGAAATTAAGACATTAATTAATTCAAACATTTGGAAGAGCCTTT 662
 2281 ACATCAGAACACTGGCTGTTAGGATATATAAAGTAAAGCACCCTGATTAACAGGAGACA 2340


```
PI KOJI HAYASHI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68, C12P21/08, C12N15/00, C12N5/00 CC
Secretory protein or membrane protein
FH Key Location/Qualifiers
FT CDS (94), (951).
FEATURES
    source
        1..2510
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
ORIGIN
    Query Match 38.5%; Score 954; DB 6; Length 2510;
    Best Local Similarity 100.0%; Pred. No. 0;
    Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1528 TTGTATGATAGGCGAGTAAGTGAGGAAACATGCAACATGAAACAGGAAACCTGAAGAG 1587
DB 1 TTGTATGATAGGCGAGTAAGTGAGGAAACATGCAACATGAAACAGGAAACCTGAAGAG 60
QY 1588 GGATTAGCCCTAATATAAAGCATTGTCAACATGTCATGCTGATGCTATTGATGATG 1647
DB 61 GGATTAGCCCTAATATAAAGCATTGTCAACATGTCATGCTGATGCTATTGATGATG 120
QY 1648 TTTGCTGTCCACTGTAACCTGGGTCAAGCAATGCTATCTAGTCCAAAGTGTAGTCTG 1707
DB 121 TTTGCTGTCCACTGTAACCTGGGTCAAGCAATGCTATCTAGTCCAAAGTGTAGTCTG 180
QY 1708 GCCTCATACATATGATGTCACAGCAATATCTTAGATGATTTAGAGAGCTTACTTT 1767
DB 181 GCCTCATACATATGATGTCACAGCAATATCTTAGATGATTTAGAGAGCTTACTTT 240
QY 1768 TGGCTAAGGCAAAATACAGATGAACATGCAAGTAACTCTTGTGGGATTTATGCTAT 1827
DB 241 TGGCTAAGGCAAAATACAGATGAACATGCAAGTAACTCTTGTGGGATTTATGCTAT 300
QY 1828 CAGATAGCTGAATGGCTAATAGAACTACCTGGTGGTAATAACACCTGGAAATACAGC 1887
DB 301 CAGATAGCTGAATGGCTAATAGAACTACCTGGTGGTAATAACACCTGGAAATACAGC 360
QY 1888 CACATAGCACTGGTGGGAAAGTATGCTCTTCTTAATGAACAGCAGCCTATATAATCATG 1947
DB 361 CACATAGCACTGGTGGGAAAGTATGCTCTTCTTAATGAACAGCAGCCTATATAATCATG 420
QY 1948 AGGACTCTAGATGATAGATTTGTTTGGTATTTTGGAGGGTATTGGCTATTCTGCT 2007
DB 421 AGGACTCTAGATGATAGATTTGTTTGGTATTTTGGAGGGTATTGGCTATTCTGCT 480
QY 2008 GATGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAGGAGAAATCCCAAGAC 2067
DB 481 GATGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAGGAGAAATCCCAAGAC 540
QY 2068 ATTCGGGAAAGTGACTATTTTACCCACAGGGAGATTCCTGTGTAGACAAAGCAGATCC 2127
DB 541 ATTCGGGAAAGTGACTATTTTACCCACAGGGAGATTCCTGTGTAGACAAAGCAGATCC 600
QY 2128 CCTACTTTGTTGAATTCCTTATGTTAATAATGCTATACAGATTTGGAGAAATGAG 2187
DB 601 CCTACTTTGTTGAATTCCTTATGTTAATAATGCTATACAGATTTGGAGAAATGAG 660
QY 2188 CTGGATTTTCGTACACCCCGAGTTTTCACCGAACACGTAATGCTGAGATTCGAATTAAG 2247
DB 661 CTGGATTTTCGTACACCCCGAGTTTTCACCGAACACGTAATGCTGAGATTCGAATTAAG 720
QY 2248 GACATTTAAATCAAACTTTGGAGAGAGCTTTTACATCAGACACTGCTGCTTTAGGATA 2307
DB 721 GACATTTAAATCAAACTTTGGAGAGAGCTTTTACATCAGACACTGCTGCTTTAGGATA 780
QY 2308 TATAAGTAAAGACCTGTATAACAGGGAGACATTAGATCAACAACCTCGAGTCCACCAAC 2367
DB 1 TTGTATGATAGGCGAGTAAGTGAGGAAACATGCAACATGAAACAGGAAACCTGAAGAG 60

Db 781 TATAAGTAAAGACCTGTATAACAGGGAGACATTAGATCAACAACCTCGAGTCCACCAAC 840
QY 2368 ATTTTCCCAAAACAGAGATTTTGTCAAGAGAGACTACCAAAAGAGCGTGGTACATT 2427
DB 841 ATTTTCCCAAAACAGAGATTTTGTCAAGAGAGACTACCAAAAGAGCGTGGTACATT 900
QY 2428 AAAAAATAAGCTGTTTTTAAAGAAAGCGAAGAAATATCTTAAGAGAGACTGTTTAA 2481
DB 901 AAAAAATAAGCTGTTTTTAAAGAAAGCGAAGAAATATCTTAAGAGAGACTGTTTAA 954

RESULT 10
AK075380 2510 bp mRNA linear PRI 03-SEP-2002
LOCUS Homo sapiens cDNA PSEC0070 fis, clone NT2RP2001508, moderately
DEFINITION similar to OLIGONACCHARYL TRANSFERASE STT3 SUBUNIT.
ACCESSION AK075380.1 GI:22761428
VERSION AK075380.1
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayashi, K.,
Ishii, S., Saiko, K., Yamamoto, J., Wakamatsu, A., Nagai, T.,
Nakamura, Y., Nagahara, K., Sugano, S. and Isogai, T.
TITLE HRI human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2510)
AUTHORS Isogai, T. and Yamamoto, J.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
[ E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986 ]
COMMENT HRI human cDNA sequencing project; cDNA 5' - 3' -end one pass
sequencing, clone selection and full insert sequencing; Helix
Research Institute (supported by Japan Key Technology Center etc.);
cDNA library construction; Institute of Medical Science, University
of Tokyo, Laboratory of Genome Structure, Human Genome Center.
FEATURES
    source
        1..2510
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="NT2RP2001508"
            /cell_line="NT2"
            /cell_type="teratocarcinoma"
            /clone_lib="NT2R22"
            /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
            precursor cells after 2-weeks retinoic acid (RA)
            induction"
            94..954
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="BAC11581.1"
            /db_xref="GI:22761428"
            /translation="WLMMLMLNFAVHCTWTWTSNAYSSPSVVLASVYHNDGTRILDDF
            REAYFWLQNDIEHARVSWMDYGYQAGMANRTLLVDNNNNWNNSHIALVKWAMSSNE
            TAAVKIMETLDDVYLVIFGVIGVGGDDINKFLMMVRIAGEBHPKIDRESYPTQG
            EFRVKGAGSLTINCLMKNKYRFGEMQLDFTPTGDFRTRNAIGNKDKIKFHLBE
            AFTSHWLVRIVYKVKAPDNRTLDHPRVTNIPPKYLSKTKTKRGRYIKNKLVPK
            KGKTKISKTV"
ORIGIN
    Query Match 38.5%; Score 954; DB 9; Length 2510;
    Best Local Similarity 100.0%; Pred. No. 0;
    Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1528 TTGTATGATAGGCGAGTAAGTGAGGAAACATGCAACATGAAACAGGAAACCTGAAGAG 1587
DB 1 TTGTATGATAGGCGAGTAAGTGAGGAAACATGCAACATGAAACAGGAAACCTGAAGAG 60
```

```

Db      721  GCTGAGATTGGAATTAAGGACATTAATTCACCAATTTGGAAGAAGCCCTTTACATCAGAA 780
Qy      2290  CACTGGCTTTAGGATATATAAGTAAGTAAGCACTGATAACAGGAGACATTTAGATCAC 2349
Db      781  CACTGGCTTTAGGATATATAAGTAAGTAAGCACTGATAACAGGAGACATTTAGATCAC 840
Qy      2350  AAACCTCGAGTCACCAACATTTCCCAAAACAGAGATTTTCTCAAAAGACATACCAAA 2409
Db      841  AAACCTCGAGTCACCAACATTTCCCAAAACAGAGATTTTCTCAAAAGACATACCAAA 900
Qy      2410  AGGAAGCGTGGCTACATTAATAAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAG 2469
Db      901  AGGAAGCGTGGCTACATTAATAAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAG 960
Qy      2470  AAGACTGTTTAA 2481
Db      961  AAGACTGTTTAA 972

RESULT 8
AX136135
LOCUS      2510 bp      DNA      linear      PAT 30-MAY-2001
DEFINITION Sequence 57 from Patent EP1067182.
ACCESSION AX136135
VERSION    AX136135.1 GI:14272543
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1
AUTHORS    Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and Hayashi, K.
TITLE      Secretory protein or membrane protein
JOURNAL    Patent: EP 1067182-A 57 10-JAN-2001; Helix Research Institute (JP)
FEATURES   source
            1..2510
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAC39710.1"
            /db_xref="GI:14272543"
            /translation="MIMLMMLMFVHCTWTSNAYSSPSVVLASYNHGDGRNILDPE
            REAYFVLRONTDEHAFVMSWDYGYQIAGMANRTLVNDNTMNSHIALVKKMSSNE
            TAAKIMRLDLDVYLVIEGGVIGSGDDINKFLMWVRLAEGHPKIRSDYFTPG
            EFRVDKAGSFTILNCYKMSYRFGSEWOLFRTPPGDFTRNNAIGNKOIKPKHLEE
            AFTSEHWLVRIYKVPADPNRETLDHKPRVTNIFPKQKYLKSKTKTKRKGVIKKNLVPK
            KGKKISKKTIV"

CDS
Query Match 38.5%; Score 954; DB 6; Length 2510;
Best Local Similarity 100.0%; Pred No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1528  TTGTATGATAGGCAGGTAAAGTGAAGAAACATGCACTGAACAGGAAAAAAGTGAAGAG 1587
Db      1  TTGTATGATAGGCAGGTAAAGTGAAGAAACATGCACTGAACAGGAAAAAAGTGAAGAG 60
Qy      1588  GGATTAGGCCTTAATATAAAGCATTTGACCATGTTGATGCTGCTATGATGATGATG 1647
Db      61  GGATTAGGCCTTAATATAAAGCATTTGACCATGTTGATGCTGCTATGATGATGATG 120
Qy      1648  TTGTGCTGCACATGTTACCTGGGTCAACAAGATGCTTACTAGTCCAAAGTGTAGTCTG 1707
Db      121  TTGTGCTGCACATGTTACCTGGGTCAACAAGATGCTTACTAGTCCAAAGTGTAGTCTG 180
Qy      1708  GCCTCATACATCATGATGTCACAGCAATATCTTAGATGATTTTATAGAAAGCTTACTTT 1767
Db      181  GCCTCATACATCATGATGTCACAGCAATATCTTAGATGATTTTATAGAAAGCTTACTTT 240

```

```

Qy      1769  TGGCTAAGGCAAAATACAGATGAACATGCACAGTAAATGTCTTGGTGGGATTTATGGCTAT 1827
Db      241  TGGCTAAGGCAAAATACAGATGAACATGCACAGTAAATGTCTTGGTGGGATTTATGGCTAT 300
Qy      1828  CAGATAGCTGGAAATGGCTAAATAGAACTACGTTGGTGGTAAATAACACCTCGAAATAACAGC 1887
Db      301  CAGATAGCTGGAAATGGCTAAATAGAACTACGTTGGTGGTAAATAACACCTCGAAATAACAGC 360
Qy      1889  CACATAGCACTGGTGGGAAAGCTATGTCTTCTTAATGAAACAGCAGCCTATAAAATCATG 1947
Db      361  CACATAGCACTGGTGGGAAAGCTATGTCTTCTTAATGAAACAGCAGCCTATAAAATCATG 420
Qy      1948  AGGACTCTAGATGCTAGATTAATGTTTGGTATTTTGGAGGGGTATTGCTGCTATTTCTGGT 2007
Db      421  AGGACTCTAGATGCTAGATTAATGTTTGGTATTTTGGAGGGGTATTGCTGCTATTTCTGGT 480
Qy      2008  GATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAGGAGAACATCCCAAGAC 2067
Db      481  GATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAGGAGAACATCCCAAGAC 540
Qy      2068  ATTCGGGAAAGTGACTATTTTACCCCAAGGAGAAATCCGTTGTAGACAAAGCAGGATCC 2127
Db      541  ATTCGGGAAAGTGACTATTTTACCCCAAGGAGAAATCCGTTGTAGACAAAGCAGGATCC 600
Qy      2128  CCTACTTTGTTGAATGGCTTATGTATAAATGTCATCTACTACAGATTTGGAGAAATGCAG 2187
Db      601  CCTACTTTGTTGAATGGCTTATGTATAAATGTCATCTACTACAGATTTGGAGAAATGCAG 660
Qy      2188  CTGGATTTTCTGTACACCCCAAGGTTTTCACCGAACACGTAATGCTGAGATTTGAAAATAAG 2247
Db      661  CTGGATTTTCTGTACACCCCAAGGTTTTCACCGAACACGTAATGCTGAGATTTGAAAATAAG 720
Qy      2248  GACATTAATTCACAAATTTGGAAGAAGCCCTTTACATCAGAAACACTGGCTTTGTAGATA 2307
Db      721  GACATTAATTCACAAATTTGGAAGAAGCCCTTTACATCAGAAACACTGGCTTTGTAGATA 780
Qy      2308  TATAAAGTAAAAGCACCTGTATAACAGGAGACATTTAGATCACAACCTCGAGTCACCAAC 2367
Db      781  TATAAAGTAAAAGCACCTGTATAACAGGAGACATTTAGATCACAACCTCGAGTCACCAAC 840
Qy      2368  ATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAAGGAGAGCGTGGCTACATT 2427
Db      841  ATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAAGGAGAGCGTGGCTACATT 900
Qy      2428  AAAAAATAAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAGAAAGACTGTTTTAA 2481
Db      901  AAAAAATAAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAGAAAGACTGTTTTAA 954

RESULT 9
BD123520
LOCUS      2510 bp      DNA      linear      PAT 18-SEP-2002
DEFINITION Secretory protein or membrane protein.
ACCESSION BD123520
VERSION    BD123520.1 GI:23218465
KEYWORDS   JP 2002017376-A/29.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 2510)
AUTHORS    Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and Hayashi, K.
TITLE      Secretory protein or membrane protein
JOURNAL    Patent: JP 2002017376-A 29 22-JAN-2002; HELIX RESEARCH INSTITUTE
COMMENT    OS Homo sapiens (human)
            PN JP 2002017376-A/29
            PD 22-JAN-2002
            PF 07-JUL-2000 JP 2000253173
            PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, TOMOYASU
            PI SUGIYAMA,

```

QY 1690 AGTCCAAAGTGTAGCTCGGCTCATACATCATGATGGACACAGGAATATCTTAGATGAT 1749
 Db 181 AGTCCAAAGTGTAGCTCGGCTCATACATCATGATGGACACAGGAATATCTTAGATGAT 240
 QY 1750 TTTAGAGAGCTTACTTTTGGCTAAGCCAAATACAGATGAACATGCACGAGTAATGTCT 1809
 Db 241 TTTAGAGAGCTTACTTTTGGCTAAGCCAAATACAGATGAACATGCACGAGTAATGTCT 300
 QY 1810 TGGTGGGATTTAGCTATCATAGATGCTGATGCTAATAGAACATCGTTGGTGAATAT 1869
 Db 301 TGGTGGGATTTAGCTATCATAGATGCTGATGCTAATAGAACATCGTTGGTGAATAT 360
 QY 1870 AACACCTGGATTAACACCCATAGACATGCTGGTGGGAAAAGCTATGCTTCTAATGAAACA 1929
 Db 361 AACACCTGGATTAACACCCATAGACATGCTGGTGGGAAAAGCTATGCTTCTAATGAAACA 420
 QY 1930 GCAGCTTAAATCATAGAGATCTAGATGATGATTAATGTTTGGTATTTTGGAGGG 1989
 Db 421 GCAGCTTAAATCATAGAGATCTAGATGATGATTAATGTTTGGTATTTTGGAGGG 480
 QY 1990 GTTATTGGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAA 2049
 Db 481 GTTATTGGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAA 540
 QY 2050 GGAGACATCCCAAGACATTCGGGAAGTACTATTTTACCCACAGGAGAAATTCCTGT 2109
 Db 541 GGAGACATCCCAAGACATTCGGGAAGTACTATTTTACCCACAGGAGAAATTCCTGT 600
 QY 2110 GTAGACAAAGCAGGATCCCTACTTTTGGTGAATGATGATTAATAAATGTCTACTATC 2169
 Db 601 GTAGACAAAGCAGGATCCCTACTTTTGGTGAATGATGATTAATAAATGTCTACTATC 660
 QY 2170 AGATTGGAGAAATCGAGTGGATTTTCTGACACCCAGGTTTGGACGACAGTAAT 2229
 Db 661 AGATTGGAGAAATCGAGTGGATTTTCTGACACCCAGGTTTGGACGACAGTAAT 720
 QY 2230 GCTGAGATTGGAATTAAGACATTAATAATCAACATTTTGGAAAGAGCTTTTACATCAGAA 2289
 Db 721 GCTGAGATTGGAATTAAGACATTAATAATCAACATTTTGGAAAGAGCTTTTACATCAGAA 780
 QY 2290 CACTGGCTCTTAGGATATTAAGTAAGACCTGATAACAGGAGACATTAGATCAC 2349
 Db 781 CACTGGCTCTTAGGATATTAAGTAAGACCTGATAACAGGAGACATTAGATCAC 840
 QY 2350 AAACCTCGAGTACCACATTTTCCCAAAACAGAGTATTTGTCAAAGAGACTACCAA 2409
 Db 841 AAACCTCGAGTACCACATTTTCCCAAAACAGAGTATTTGTCAAAGAGACTACCAA 900
 QY 2410 AGGAAGCTGGCTACATTAATAATAGCTGGTTTAAAGAGGCAAGAAATATCTAAG 2469
 Db 901 AGGAAGCTGGCTACATTAATAATAGCTGGTTTAAAGAGGCAAGAAATATCTAAG 2529
 QY 2470 AAGACTGTTAA 2481
 Db 961 AAGACTGTTAA 972

RESULT 7
 LOCUS B063986 2546 bp DNA linear PAT 27-AUG-2002
 DEFINITION Secreted proteins and polynucleotides encoding them.
 ACCESSION B063986
 VERSION B063986.1 G1:22609589
 KEYWORDS JP 2001506848-A/10.
 SOURCE unidentified
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 2546)
 AUTHORS Jacobs,K., McCoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D., Treacy,M., Spaulding,V. and Agostino,M.J.
 TITLE Secreted proteins and polynucleotides encoding them
 JOURNAL Patent: JP 2001506848-A 10 29-MAY-2001;

GENETICS INSTITUTE INC
 PN JP 2001506848-A/10
 PD 29-MAY-2001
 PF 12-DEC-1997 JP 1998525996
 PR 13-DEC-1996 US 08/766263
 PI KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE, PI
 DAVID MERBERG.
 PI MAURICE TREACY,VIKKI SPAULDING,MICHAEL J AGOSTINO PC
 C12N15/12,C12N5/10,C07K14/47,A61K38/17
 CC Strandedness: Double;
 CC Topology: Linear;
 FH Key Location/Qualifiers.
 FEATURES
 source
 1..2546
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 ORIGIN
 Query Match 39.2%; Score 972; DB 6; Length 2546;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1510 ABAAGAAACCAAGGAAATTTCTATGATAGGACAGGTAAAGTGAGAAACATGCACTGAA 1569
 Db 1 ABAAGAAACCAAGGAAATTTCTATGATAGGACAGGTAAAGTGAGAAACATGCACTGAA 60
 QY 1570 CAGGAAAAAATCGAAGAGGATTTAGGCCCTAATATAAAGCATTGTCCACCATTTGATG 1629
 Db 61 CAGGAAAAAATCGAAGAGGATTTAGGCCCTAATATAAAGCATTGTCCACCATTTGATG 120
 QY 1630 CTGATGCTATTGATGATGTTTCTGCTCCACCTGATGCTGGGTCAACGATGCTACTCT 1689
 Db 121 CTGATGCTATTGATGATGTTTCTGCTCCACCTGATGCTGGGTCAACGATGCTACTCT 180
 QY 1690 AGTCCAAAGTGTAGTCTCTGCTCATACATCATGATGSCACAGGAAATCTTAGATGAT 1749
 Db 181 AGTCCAAAGTGTAGTCTCTGCTCATACATCATGATGSCACAGGAAATCTTAGATGAT 240
 QY 1750 TTTAGAGAGCTTACTTTTGGCTAAGCAAAATACAGATGAACATGACGAGTAATGTCT 1809
 Db 241 TTTAGAGAGCTTACTTTTGGCTAAGCAAAATACAGATGAACATGACGAGTAATGTCT 300
 QY 1810 TGGTGGGATTTAGCTATCAGATAGCTGGAATGGCTAATAGAACTAGTGGTGGATAAT 1869
 Db 301 TGGTGGGATTTAGCTATCAGATAGCTGGAATGGCTAATAGAACTAGTGGTGGATAAT 360
 QY 1870 AACACCTGGAATAACAGACCATAGCATGCTGGTGGGAAAAGCTATGCTTCTAATGAAACA 1929
 Db 361 AACACCTGGAATAACAGACCATAGCATGCTGGTGGGAAAAGCTATGCTTCTAATGAAACA 420
 QY 1930 GCAGCTTAAATCATAGGACTCTAGATGATGATTAATGTTTGGTATTTTGGAGGG 1989
 Db 421 GCAGCTTAAATCATAGGACTCTAGATGATGATTAATGTTTGGTATTTTGGAGGG 480
 QY 1990 GTTATTGGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAA 2049
 Db 481 GTTATTGGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAA 540
 QY 2050 GGAGAACATCCCAAGACATTCGGGAAGTACTATTTTACCCACAGGAGAAATTCCTGT 2109
 Db 541 GGAGAACATCCCAAGACATTCGGGAAGTACTATTTTACCCACAGGAGAAATTCCTGT 600
 QY 2110 GTAGACAAAGCAGGATCCCTACTTTTGGTGAATGATGATTAATAAATGTCTACTATC 2169
 Db 601 GTAGACAAAGCAGGATCCCTACTTTTGGTGAATGATGATTAATAAATGTCTACTATC 660
 QY 2170 AGATTGGAGAAATCGAGTGGATTTTCTGACACCCAGGTTTGGACGACAGTAAT 2229
 Db 661 AGATTGGAGAAATCGAGTGGATTTTCTGACACCCAGGTTTGGACGACAGTAAT 720
 QY 2230 GCTGAGATTGGAATTAAGACATTAATAATCAACATTTTGGAAAGAGCTTTTACATCAGAA 2289

/translation="WAPSPAPSKSKSLSNSPSGLMALNSRHHGHPGAQCAHKA
AGAAPKPPAPAGLGGSSQAPAGWOSLLSFTLLFLAMLAGFSRLFAVIRPESLHIF
DPWNYRSTHILASHGFYBLNWFDERAWYPLGRIVGGTVVYPMITAGLHMLNLT
NITVHRDVCFLTSLFSLTISITFLRLNQGAGLLAACFIATVPVIGRSVAG
SPNEGIAIFALQFTYLVWVKSVKSGVFTWCCCLSYFVWVSWAGGVFLINILPH
FVLLLMORYSEVITAYSTFVIGLILSMOIPRVGPOPIRTSEHMAAGVALLQAY
AFLOYLRDLTQBEQTLFELGSLAAGAVELSVLYLTGTGYIAPMSGRVSLWDTGY
AKIHIFIIASVSEHQTTWVSFFDLHLVLCFFPAGLWFCIKNINDERVGA"

ORIGIN

Query Match 51.4%; Score 1276; DB 9; Length 1664;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1326; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGGAGCCCTCGGCCCGGAGACAGCACTGCTCCCTCAACTGCTCCCGTGG 60
DB 28 ATGCGGAGCCCTCGGCCCGGAGACAGCACTGCTCCCTCAACTGCTCCCGTGG 87
QY 61 AGTGGCCCTATGCGCCCTGGGAAACAGCCGCGACCGCCACCGGCCCGGCGCCAGTGC 120
DB 88 AGTGGCCCTATGCGCCCTGGGAAACAGCCGCGACCGCCACCGGCCCGGCGCCAGTGC 147
QY 121 GCGCACAGCG 180
DB 148 GCGCACAGCG 207
QY 181 GGGCTGTCGAGCG 240
DB 208 GGGCTGTCGAGCG 267
QY 241 TGGCTTCCGCGCTACGTCGCGCTTTCGCGGTCATCCGCTTCGAAAGCATCATCCAC 300
DB 268 TGGCTTCCGCGCTACGTCGCGCTTTCGCGGTCATCCGCTTTCGAAAGCATCATCCAC 327
QY 301 GAGTTCACCCCGTGGTTTAACTATAGATCAACACATCATCTTTCGATCTCATGGGTTCTAT 360
DB 328 GAGTTCACCCCGTGGTTTAACTATAGATCAACACATCATCTTTCGATCTCATGGGTTCTAT 387
QY 361 GAATTTTAAATGGTTTGAAGAGCATGGTATCCACTAGGAGATAGTAGGTTGGT 420
DB 388 GAATTTTAAATGGTTTGAAGAGCATGGTATCCACTAGGAGATAGTAGGTTGGT 447
QY 421 ACTGTTTACCCAGGTTGATGATPACCGCTGGCCCTTATTCATTTGAAATTAATACATTG 480
DB 448 ACTGTTTACCCAGGTTGATGATPACCGCTGGCCCTTATTCATTTGAAATTAATACATTG 507
QY 481 AACATACCTTTCACATAGAGACGTATGTTGTTCCCTTGCACCACTTTTAGCGGCTT 540
DB 508 AACATACTGTTTCACATAGAGACGTATGTTGTTCCCTTGCACCACTTTTAGCGGCTT 567
QY 541 ACATCTATCTACTTTCCTGCTTACAAGAGAACTTTGGAACCAAGGACGAGGACTTTTA 600
DB 568 ACATCTATCTACTTTCCTGCTTACAAGAGAACTTTGGAACCAAGGACGAGGACTTTTA 627
QY 601 GCTGCTTGTGTTTATGCTATGTTACAGAGCTACATATCCGCTAGTAGCTGGATCCCTT 660
DB 628 GCTGCTTGTGTTTATGCTATGTTACAGAGCTACATATCCGCTAGTAGCTGGATCCCTT 687
QY 661 GATAATGAAGGCACTGCTATTTTTCGACTTCAGTTCACATATTTATGGTAAATPCT 720
DB 688 GATAATGAAGGCACTGCTATTTTTCGACTTCAGTTCACATATTTATGGTAAATPCT 747
QY 721 GTAAAACTGGGTCAGTTTTTGGACAATGTGCTGCTTATCCATTTCTATATGTC 780
DB 748 GTAAAACTGGGTCAGTTTTTGGACAATGTGCTGCTTATCCATTTCTATATGTC 807
QY 781 TCTGCTTGGGTTGTTATGTTATTTATCATCTTTCATCTCATCTATTTTGTGTTG 840
DB 808 TCTGCTTGGGTTGTTATGTTATTTATCATCTTTCATCTCATCTATTTTGTGTTG 867
QY 841 TTACTGATGAGAGATACACAAAGAGCTCATAGCATATAGCACCTTCTACATCTG 900
DB 868 TTACTGATGAGAGATACACAAAGAGCTCATAGCATATAGCACCTTCTACATCTG 927

QY 901 GGTTTAATATTATCAATGCAGATACCTTTTGTGGGATTCAGCCATCAGAACAGTGAA 960
DB 928 GGTTTAATATTATCAATGCAGATACCTTTTGTGGGATTCAGCCATCAGAACAGTGAA 987
QY 961 CACATGGCAGCTGCAAGGTGCTTTTGCATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1020
DB 988 CACATGGCAGCTGCAAGGTGCTTTTGCATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1047
QY 1021 AGAGACCGATTAAACAAACAGAGTTCCAGACCCCTTTCTTTTGGGTGATCAGTAGCT 1080
DB 1048 AGAGACCGATTAAACAAACAGAGTTCCAGACCCCTTTCTTTTGGGTGATCAGTAGCT 1107
QY 1081 GCAGGTGCTGTTGTTTCCCTTAGTGTGTCATCTATTGACTTATACAGGTTACATGACCATTG 1140
DB 1108 GCAGGTGCTGTTGTTTCCCTTAGTGTGTCATCTATTGACTTATACAGGTTACATGACCATTG 1167
QY 1141 AGTGCAGGTTTATCTATTGTTGGGATCTGGGTATGCAAAATATACATTCATTAAT 1200
DB 1168 AGTGCAGGTTTATCTATTGTTGGGATCTGGGTATGCAAAATATACATTCATTAAT 1227
QY 1201 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTTTGTGATCTACATATT 1260
DB 1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTTTGTGATCTACATATT 1287
QY 1261 CTTGTATGATCTTCCGAGAGGCTTGGTTCTGTCATCAAAATATCAACGATGAAAGA 1320
DB 1288 CTTGTATGATCTTCCGAGAGGCTTGGTTCTGTCATCAAAATATCAACGATGAAAGA 1347
QY 1321 GTATTG 1327
DB 1348 GTATTG 1354

RESULT 6
AX099510
LOCUS 2546 bp DNA linear PAT 02-APR-2001
DEFINITION Sequence 150 from Patent WO0119988.
ACCESSION AX099510
VERSION AX099510.1 GI:13538588
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Jacobs K., McCoy J.M., Lavallie E.R., Collins-Racie L.A., Evans C., Merberg D., Treacy M., Bowman M.R., Spaulding V. and Agostino M.J.
TITLE Secreted proteins and polynucleotides encoding them
JOURNAL Patent: WO 0119988-A 150 22-MAR-2001;
Genetics Institute, Inc. (US)
FEATURES
source
1. 2546
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

Query Match 39.2%; Score 972; DB 6; Length 2546;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1510 AAAAGAAACCAAGGAAATTTGTATGATAAGGCAGGTAAAGTAGGAGAAACATGCAACTGAA 1569
DB 1 AAAAGAAACCAAGGAAATTTGTATGATAAGGCAGGTAAAGTAGGAGAAACATGCAACTGAA 60
QY 1570 CAGGAAAAACATGAAGAGGATTAGGCCCTTAATAAAAAAGCATTTGCACCATTTGATG 1629
DB 61 CAGGAAAAACATGAAGAGGATTAGGCCCTTAATAAAAAAGCATTTGCACCATTTGATG 120
QY 1630 CTGATGCTATTGATGATGTTTGTGCTGCTCCACTGTACCTGGGTCAACAAGCAATGCTACTCT 1689
DB 121 CTGATGCTATTGATGATGTTTGTGCTGCTCCACTGTACCTGGGTCAACAAGCAATGCTACTCT 180

```
Query Match 51.4%; Score 1276; DB 6; Length 1664;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1326; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGGAGCCCTCGGCCCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGG 60
DB 28 ATGCGGAGCCCTCGGCCCGGAGAGCAAGTCGTCCTCAACTCGTCCCGG 87
QY 61 AGTGGCTATGCGCCTGGAAGAACAGCGGSCACGCGGCAACGCGGCGCCGAGTGC 120
DB 88 AGTGGCTATGCGCCTGGAAGAACAGCGGSCACGCGGCAACGCGGCGCCGAGTGC 147
QY 121 GCGCACAGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 180
DB 148 GCGCACAGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 207
QY 181 GGGCTGTGCGAGCGCGCTGGGTGCGAGTGCCTTCTCTCTTCAACATCTCTTCTGCCC 240
DB 208 GGGCTGTGCGAGCGCGCTGGGTGCGAGTGCCTTCTCTCTTCAACATCTCTTCTGCCC 267
QY 241 TGGCTTGGCGGCTTCAGCTCGCGCCTCTTCCGCGTGCATCCGCTTCGAAGCATCATCCAC 300
DB 268 TGGCTTGGCGGCTTCAGCTCGCGCCTCTTCCGCGTGCATCCGCTTCGAAGCATCATCCAC 327
QY 301 GAGTTCGACCGCTGGTTAACTTAACTAGATCAACATCATCTTCCATCTCATCGTCTAT 360
DB 328 GAGTTCGACCGCTGGTTAACTTAACTAGATCAACATCATCTTCCATCTCATCGTCTAT 387
QY 361 GAAATTTTAAATTTGGTTTGAAGAGAGCATGTATCCACTAGGAGAAATAGTAGTGGT 420
DB 388 GAAATTTTAAATTTGGTTTGAAGAGAGCATGTATCCACTAGGAGAAATAGTAGTGGT 447
QY 421 ACTGTTTACCCAGGTTGATGATTAACCGCTGGCGCTTATTCATGATGATTTAAATACATG 480
DB 448 ACTGTTTACCCAGGTTGATGATTAACCGCTGGCGCTTATTCATGATGATTTAAATACATG 507
QY 481 AACATAACTGTTCATAGAGAGCATGTATGCTTCTTGCACCAACTTTTACGGCGCTT 540
DB 508 AACATAACTGTTCATAGAGAGCATGTATGCTTCTTGCACCAACTTTTACGGCGCTT 567
QY 541 ACATCTATATCTTCTTCTGCTTACAGAGAGACTTTGGAACCAAGGAGCAGACTTTTA 600
DB 568 ACATCTATATCTTCTTCTGCTTACAGAGAGACTTTGGAACCAAGGAGCAGACTTTTA 627
QY 601 GCTGCTGTTTTATGCTATTTTACAGGCTACATATCGGTCAGTACGTCGATCTTT 660
DB 628 GCTGCTGTTTTATGCTATTTTACAGGCTACATATCGGTCAGTACGTCGATCTTT 687
QY 661 GATATGAGGCACTGCTATTTTGCACCTTCAGTTCACATCTATTTATGGTAAATCT 720
DB 688 GATATGAGGCACTGCTATTTTGCACCTTCAGTTCACATCTATTTATGGTAAATCT 747
QY 721 GTAAAACTGGGTGAGTATTTTGAACAATGTGCTGCTATTCCTATTTCTATATGCTC 780
DB 748 GTAAAACTGGGTGAGTATTTTGAACAATGTGCTGCTATTCCTATTTCTATATGCTC 807
QY 781 TCTGCTGGGTGGTATGATTTATTCATCAATCTTATTCCTACGTCATGATTTGCTG 840
DB 808 TCTGCTGGGTGGTATGATTTATTCATCAATCTTATTCCTACGTCATGATTTGCTG 867
QY 841 TTACTGATGAGAGATACAGCAAGAGTCTACATAGCATATAGCACTTTCTACATTTG 900
DB 868 TTACTGATGAGAGATACAGCAAGAGTCTACATAGCATATAGCACTTTCTACATTTG 927
QY 901 GGTTTATATTTATCAATGAGATACCTTTTGTGGGATTCAGCAATCAGAACAGTGA 960
DB 928 GGTTTATATTTATCAATGAGATACCTTTTGTGGGATTCAGCAATCAGAACAGTGA 987
QY 961 CACATGGAGCTGAGGTGCTTTGCAATTCCTGCAAGCTATGCTTTCTTGAGTATCTG 1020
DB 988 CACATGGAGCTGAGGTGCTTTGCAATTCCTGCAAGCTATGCTTTCTTGAGTATCTG 1047
QY 1021 AGAGACCGATTAACAAAAACAGAGTTCAGACCCCTTTTCTTTTGGGTGTCATCACTAGCT 1080

1048 AGAGACCGATTAACAAAAACAGAGTTCAGACCCCTTTTCTTTTGGGTGTCATCACTAGCT 1107
1081 GCAGGTGCTGCTGCTTCTTCTAGTGTCTATCTATTGACTTATACAGGTACATGTCACCATGG 1140
1108 GCAGGTGCTGCTGCTTCTTCTAGTGTCTATCTATTGACTTATACAGGTACATGTCACCATGG 1167
1141 AGTGCAGGTTTTATTCTATTGTTGGATCTGGGTATGCGGATCAAAAATACATCTCAATATT 1200
1168 AGTGCAGGTTTTATTCTATTGTTGGATCTGGGTATGCGGATCAAAAATACATCTCAATATT 1227
1201 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTGTGATCTACATATT 1260
1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTGTGATCTACATATT 1287
1261 CTTGTATGATCTTCCAGCAGGCTTGTGTCATCAAAAATATCAACGATGAAGA 1320
1288 CTTGTATGATCTTCCAGCAGGCTTGTGTCATCAAAAATATCAACGATGAAGA 1347
1321 GTATTGG 1327
1348 GTATTGG 1354

RESULT 5
AK074587 1664 bp mRNA linear PRI 03-SEP-2002
LOCUS Homo sapiens cDNA FLJ90106 f1s, clone HEMBA1006430, weakly similar
DEFINITION to Human putative transmembrane protein precursor (B5) mRNA.
ACCESSION AK074587
VERSION AK074587.1 GI:22760122
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
Kojima,S., Nagahara,K., Masuho,Y., Ono,T., Okano,K., Yoshikawa,Y.,
Aotaka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
Ninomiya,K.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 1664)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction:
Institute of Medical Science, University of Tokyo, Laboratory of
Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
sequencing and clone selection: Helix Research Institute (supported
by Japan Key Technology Center etc.).
FEATURES
Location/Qualifiers
1..1664
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEMBA1006430"
/tissue_type="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/notes="cloning vector: pME18SFL3"
28..1382
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC11076.1"
/db_xref="GI:22760123"

CDS
```

DPNFNVRSTHLLASHGFYEFLLNPFDERAWYPLGRIVGGTVVPGIMITAGLHMLNTL
 NITVHRDVCVFLPTFSLTISIFLITRELMQAGGLAAACFIATVPGYIERSVAG
 SFDEGIAIFALQFTVYILWVKSVGTGVTMCCCLGSFYVMWSANGVYFIINLIPLH
 FVLLLMQVRSRVVIAVSTFVIGLILSMOIPFVGFOPRTSHMAAGVFAILOAY
 AFLOYLRDLTKQBQCTFFLGVSLAAGAVLSVLYLTGCIAPWSGRVSLWDIGY
 AKIHIFLIASVSEHQPTWVSFFDLHLVCTFFAGLWFCIKINDERVGA"

ORIGIN

Query Match 51.4%; Score 1276; DB 6; Length 1664;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1326; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCGGAGCCCTCGGCCCGGAGACACAGTGTCTCACTCACTGTCCTCCGCTGG 60
 DB |||||
 QY 28 ATGGCGGAGCCCTCGGCCCGGAGACACAGTGTCTCACTCACTGTCCTCCGCTGG 87
 DB |||||
 QY 61 AGTGGCTCATGCGCTCGGAAACAGCGCGACCGCCACCGCGGCCCGGCGCCAGTGC 120
 DB |||||
 QY 88 AGTGGCTCATGCGCTCGGAAACAGCGCGACCGCCACCGCGGCCCGGCGCCAGTGC 147
 DB |||||
 QY 121 GGCACACAGCG 180
 DB |||||
 QY 148 GCGCACAGCG 207
 DB |||||
 QY 181 GGGCTGTGCGAGCG 240
 DB |||||
 QY 208 GGGCTGTGCGAGCG 267
 DB |||||
 QY 241 TGGCTTGGCGGCTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
 DB |||||
 QY 268 TGGCTTGGCGGCTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 327
 DB |||||
 QY 301 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTTGTGATCTCATCTGCTTAT 360
 DB |||||
 QY 328 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTTGTGATCTCATCTGCTTAT 387
 DB |||||
 QY 361 GAATTTTAAATGGTTCATGAAGACGATGATGATGATGATGATGATGATGATGATGAT 420
 DB |||||
 QY 388 GAATTTTAAATGGTTCATGAAGACGATGATGATGATGATGATGATGATGATGATGAT 447
 DB |||||
 QY 421 ACTGTTTACCGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
 DB |||||
 QY 448 ACTGTTTACCGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 507
 DB |||||
 QY 481 AACATGACTGTCACATGAGACGATGATGATGATGATGATGATGATGATGATGATGATG 540
 DB |||||
 QY 508 AACATGACTGTCACATGAGACGATGATGATGATGATGATGATGATGATGATGATGATG 567
 DB |||||
 QY 541 ACATCTATATCTACTTCTCTGTTTACAGAGAACTTTGGAACCAAGGACGAGGACTTTTA 600
 DB |||||
 QY 568 ACATCTATATCTACTTCTCTGTTTACAGAGAACTTTGGAACCAAGGACGAGGACTTTTA 627
 DB |||||
 QY 601 GCTGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
 DB |||||
 QY 628 GCTGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 687
 DB |||||
 QY 661 GATAATGAAGGCACTGCTATTTTGGCACTTCAGTTTCACATATCTATTTATGGTAAATCT 720
 DB |||||
 QY 688 GATAATGAAGGCACTGCTATTTTGGCACTTCAGTTTCACATATCTATTTATGGTAAATCT 747
 DB |||||
 QY 721 GTAAARACTGGGTCAGTTTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 DB |||||
 QY 748 GTAAARACTGGGTCAGTTTGTGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 807
 DB |||||
 QY 781 TCTGCTTGGGCTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
 DB |||||
 QY 808 TCTGCTTGGGCTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 867
 DB |||||
 QY 841 TTACTGATGAGAGATACACAAAGAGTCTACATAGCATATAGCACTTTCTTACATTTGTG 900
 DB |||||
 QY 868 TTACTGATGAGAGATACACAAAGAGTCTACATAGCATATAGCACTTTCTTACATTTGTG 927
 DB |||||
 QY 901 GGTTTAATATATCAATGACATGACCTTTTGTGGGATTCAGCCAAATCAGAACAGTGAA 960
 DB |||||

DB 928 GGTTTAATATATCAATGACATACCTTTTGTGGATTCAGCAATCAGAACAGTGAA 987
 QY 961 CACATGGGAGCTGCGAGGTGCTTTTGCATGCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1020
 DB |||||
 QY 988 CACATGGGAGCTGCGAGGTGCTTTTGCATGCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1047
 DB |||||
 QY 1021 AGAGCCGATTAACAAACAGAGTTCAGACCTTTTCTTTTGGTGTATCAGTATCTGCT 1080
 DB |||||
 QY 1048 AGAGCCGATTAACAAACAGAGTTCAGACCTTTTCTTTTGGTGTATCAGTATCTGCT 1107
 DB |||||
 QY 1081 GCAGGTGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1140
 DB |||||
 QY 1108 GCAGGTGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1167
 DB |||||
 QY 1141 AGTGCAGGTTTATTAATCTTGTGGATCTGGGTATGCAAAATACATCTTCAATATT 1200
 DB |||||
 QY 1168 AGTGCAGGTTTATTAATCTTGTGGATCTGGGTATGCAAAATACATCTTCAATATT 1227
 DB |||||
 QY 1201 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTCTTCTTCTTCTTCTTCT 1260
 DB |||||
 QY 1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTCTTCTTCTTCTTCTTCT 1287
 DB |||||
 QY 1261 CTTGTATGATCTTCCAGCAGGCTTTGGTCTGATCAAAATATCAAGATGAAGA 1320
 DB |||||
 QY 1288 CTTGTATGATCTTCCAGCAGGCTTTGGTCTGATCAAAATATCAAGATGAAGA 1347
 DB |||||
 QY 1321 GTATTGT 1327
 DB |||||
 QY 1348 GTATTGT 1354
 DB |||||

RESULT 4

BD127193

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

OS

PN

PD

PF

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

PI

BD127193 1664 bp DNA linear PAT 18-SEP-2002
 Primer for synthesizing full-length cDNA and use thereof.

BD127193
 BD127193.1 GI:23222138
 JP 2002017375-A/2624.
 Homo sapiens (human)

Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 1664)
 Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
 Koga, H.

Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 2624 22-JAN-2002;
 HELIX RESEARCH INSTITUTE

OS Homo sapiens (human)
 PN JP 2002017375-A/2624
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172

PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI, HISASHI KOGA

PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
 10, C12P21/02, C12Q1/68//C12P21/08 G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof PH Key

FT CDS Location/Qualifiers
 1..1664
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

FEATURES

source

ORIGIN

Db 781 TCTGCTGGGTGGTATGATTTATCATCAATCTTATCCACTGCTATGTTGGTGG 840
QY 841 TTACTGATGACAGATACAGAAAGAGTCTCATAGCATATAGCACTTTTACATTTG 900
Db 841 TTACTGATGACAGATACAGAAAGAGTCTCATAGCATATAGCACTTTTACATTTG 900
QY 901 GGTTTAATATATCAATGACAGATACCTTTTGTGGGATTCAGCCATATCAAGCAAGTAA 960
Db 901 GGTTTAATATATCAATGACAGATACCTTTTGTGGGATTCAGCCATATCAAGCAAGTAA 960
QY 961 CACATGGAGCTGACAGTGTCTTTGCAATGCTGCAAGCTTATGCTTTCTTGCAATCTG 1020
Db 961 CACATGGAGCTGACAGTGTCTTTGCAATGCTGCAAGCTTATGCTTTCTTGCAATCTG 1020
QY 1021 AGAGACCGATTAAACAAACAAAGAGTCTCCAGACCTTTTCTTTTGGGTATCACTAGCT 1080
Db 1021 AGAGACCGATTAAACAAACAAAGAGTCTCCAGACCTTTTCTTTTGGGTATCACTAGCT 1080
QY 1081 GCAGGTGCTGTTCCTTAGTGTCATCTATTGACTTATACAGTTTACATTCGACCATGG 1140
Db 1081 GCAGGTGCTGTTCCTTAGTGTCATCTATTGACTTATACAGTTTACATTCGACCATGG 1140
QY 1141 AGTGGCAGGTTTATTCATTTGCGGATACCTGGGTATGCAAAAATACACATTCCTAATTT 1200
Db 1141 AGTGGCAGGTTTATTCATTTGCGGATACCTGGGTATGCAAAAATACACATTCCTAATTT 1200
QY 1201 GCATCAGTCTGACGATCACTACGACTGCTGGTGTCTTCTTCTTGATCTACATTT 1260
Db 1201 GCATCAGTCTGACGATCACTACGACTGCTGGTGTCTTCTTCTTGATCTACATTT 1260
QY 1261 CTTGTATGATCTTCCAGCAGGCTTTGGTCTGATCAAAAATACCAAGATGAAAGA 1320
Db 1261 CTTGTATGATCTTCCAGCAGGCTTTGGTCTGATCAAAAATACCAAGATGAAAGA 1320
QY 1321 GTATTTGCTCTATATGAAACAGTCTGCTATCTTCTGCGAGTGTATGCTGCGACTG 1380
Db 1321 GTATTTGCTCTATATGAAACAGTCTGCTATCTTCTGCGAGTGTATGCTGCGACTG 1380
QY 1381 ATGTTGATCTTGAATCCAGTCTGTGTATGCTGCTGCAATGCTTTCAAAATGTTTT 1440
Db 1381 ATGTTGATCTTGAATCCAGTCTGTGTATGCTGCTGCAATGCTTTCAAAATGTTTT 1440
QY 1441 GAGCAGTATTTGGGGATGATCATGAAAGGGAATCCACCTGTGGAGCAGCTGCT 1500
Db 1441 GAGCAGTATTTGGGGATGATCATGAAAGGGAATCCACCTGTGGAGCAGCTGCT 1500
QY 1501 GAGGATGACAAAGAAACCAAGGAAATTTCTATGATAGGCAAGTAAAGTGAAGAAACAT 1560
Db 1501 GAGGATGACAAAGAAACCAAGGAAATTTCTATGATAGGCAAGTAAAGTGAAGAAACAT 1560
QY 1561 GCAACTGAAACAGGAAACCAAGGAAATTTAGGCTTAAATATAAAGCAATTTGCTACC 1620
Db 1561 GCAACTGAAACAGGAAACCAAGGAAATTTAGGCTTAAATATAAAGCAATTTGCTACC 1620
QY 1621 ATGTTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
Db 1621 ATGTTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
QY 1681 GCCTACTCTAGTCCAAAGTGTAGTCTGCTGCTCATAATCATGATGATGATGATGATGAT 1740
Db 1681 GCCTACTCTAGTCCAAAGTGTAGTCTGCTGCTCATAATCATGATGATGATGATGATGAT 1740
QY 1741 TTAGATGATTTTAGAGAGCTTACTTTTGGTCAAGCAAAATACAGATGAACATGACGA 1800
Db 1741 TTAGATGATTTTAGAGAGCTTACTTTTGGTCAAGCAAAATACAGATGAACATGACGA 1800
QY 1801 GTATGCTCTGCTGGGATTTAGGCTATCAGATAGCTGGATGCTTAAATGAAGTACGTTG 1860
Db 1801 GTATGCTCTGCTGGGATTTAGGCTATCAGATAGCTGGATGCTTAAATGAAGTACGTTG 1860
QY 1861 GTGGATATACACTGGAATTAACAGCCATAGCTGCTGGGAAAGCTATGCTTCT 1920

Db 1861 GTGGATATTAACACCTGGAATAACAGCAATAGCACTGGTGGGAAAGCTATGCTTCT 1920
QY 1921 AATGAAACAGCAGCTATAAAATCATGAGGACTCTAGATGTAGATTAATTTTGGTTAT 1980
Db 1921 AATGAAACAGCAGCTATAAAATCATGAGGACTCTAGATGTAGATTAATTTTGGTTAT 1980
QY 1981 TTTGGAGGGGTTATTTGGGTATTTCTGGTATCAACAAATTTCTCTGGATGGTTAGG 2040
Db 1981 TTTGGAGGGGTTATTTGGGTATTTCTGGTATCAACAAATTTCTCTGGATGGTTAGG 2040
QY 2041 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAGGGA 2100
Db 2041 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAGGGA 2100
QY 2101 GAATTCCTGTTAGACAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGTATAAATG 2160
Db 2101 GAATTCCTGTTAGACAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGTATAAATG 2160
QY 2161 TCATCTACAGATTTGGGAGAAATGCGAGCTGGAATTTCTGACACCCAGGTTTGACCGA 2220
Db 2161 TCATCTACAGATTTGGGAGAAATGCGAGCTGGAATTTCTGACACCCAGGTTTGACCGA 2220
QY 2221 ACAGCTAATGCTGAGATTTGGAATAAGGACATTTAAATTTCAAAATTTGGAAGGCTTT 2280
Db 2221 ACAGCTAATGCTGAGATTTGGAATAAGGACATTTAAATTTCAAAATTTGGAAGGCTTT 2280
QY 2281 ACATCAGACACTGCTGTTAGGATATATTAAGCTTAAAGCAGCTGATACAGGAGACA 2340
Db 2281 ACATCAGACACTGCTGTTAGGATATATTAAGCTTAAAGCAGCTGATACAGGAGACA 2340
QY 2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGTATTTGTCAAGAAG 2400
Db 2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGTATTTGTCAAGAAG 2400
QY 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAAAGCAAGAA 2460
Db 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAAAGCAAGAA 2460
QY 2461 ATATCTAAGAGACTGTTTAA 2481
Db 2461 ATATCTAAGAGACTGTTTAA 2481

RESULT 3

CQ782554

LOCUS

DEFINITION

SEQUENCE

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

CQ782554 1664 bp DNA linear PAT 17-MAR-2004
Sequence 2694 from Patent EP1396543.
CQ782554
CQ782554.1 GI:45502338

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,

Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and

Koga, H.

Primer for synthesizing full length cDNA clones and their use

Patent: EP 1396543-A 2694 10-MAR-2004;

Research Association for Biotechnology (JP)

Location/Qualifiers

1..1664

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

28..1362

/note="unnamed protein product"

/codon_start=1

/protein_id="CAP85831.1"

/db_xref="GI:45502338"

/translation="MAEFPAPESKHKSLSNPWSGLMGLNRSRHGHGPGACAKHA

AGGAAPPKPAAGLSGLSPAGWQSLLSFTILFLAWLAGFSRLFAVIRFESIHEF

QY 2101 GAATTCGGTGTAGACAGAGGATCCCTACCTTTGCTTGAATTCCTTATGATAAATG 2160
 DB 2101 GAATTCGGTGTAGACAGAGGATCCCTACCTTTGCTTGAATTCCTTATGATAAATG 2160
 QY 2161 TCATACACAGATTTGAGAAATGCGAGCTGGATTTTCGTACACCCCGAGTTTGAACGA 2220
 DB 2161 TCATACACAGATTTGAGAAATGCGAGCTGGATTTTCGTACACCCCGAGTTTGAACGA 2220
 QY 2221 ACAGTAACTGTGAGATTCGAAATAGGACATTAATTCACACATTTGGAAGAGCCCTT 2280
 DB 2221 ACAGTAACTGTGAGATTCGAAATAGGACATTAATTCACACATTTGGAAGAGCCCTT 2280
 QY 2281 ACATCAGAACACTGGCTGTGTAGATATATAAGTAAAGACCTGTATAACAGGGAGACA 2340
 DB 2281 ACATCAGAACACTGGCTGTGTAGATATATAAGTAAAGACCTGTATAACAGGGAGACA 2340
 QY 2341 TTAGATCACAACCTCGAGTCACCAATTTCCCAAAACAGAGTATTTGTCAAAGAG 2400
 DB 2341 TTAGATCACAACCTCGAGTCACCAATTTCCCAAAACAGAGTATTTGTCAAAGAG 2400
 QY 2401 ACTACCAAAAGAGAGCTGCTACATTAATAATAGCTGGTTTAAAGAAAGGCAAGAAA 2460
 DB 2401 ACTACCAAAAGAGAGCTGCTACATTAATAATAGCTGGTTTAAAGAAAGGCAAGAAA 2460
 QY 2461 ATATCTAAGAGAGCTGTTTAA 2481
 DB 2461 ATATCTAAGAGAGCTGTTTAA 2481

RESULT 2
 AY074880
 LOCUS Homo sapiens source of immunodominant MHC-associated peptides
 DEFINITION (SIMP) mRNA, complete cds.
 ACCESSION AY074880
 VERSION AY074880.1 GI:19879588
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 2481)
 AUTHORS McBride, K., Baron, C., Picard, S., Martin, S., Boismenu, D., Bell, A.,
 Bergeron, J., and Perreault, C.
 TITLE The model B6dom1 minor histocompatibility antigen is encoded by a
 mouse homolog of the yeast Sit3 gene
 JOURNAL Immunogenetics 54 (8), 562-569 (2002)
 MEDLINE 22326278
 PUBMED 12439619
 REFERENCE 2 (bases 1 to 2481)
 AUTHORS McBride, K. and Perreault, C.
 TITLE Direct Submission
 JOURNAL Submitted (22-JAN-2002) Molecular Biology Group, Compatigene, 6100
 Royalmount, Montreal, QC H4P 2R2, Canada
 FEATURES
 source
 1. .2481
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 gene
 1. .2481
 /gene="SIMP"
 CDS
 1. 2481
 /gene="SIMP"
 /note="similar to Saccharomyces cerevisiae OSTT3 and Mus
 musculus protein encoded by GenBank Accession Number
 AK018758; source of minor histocompatibility antigen(s)"
 /codon_start=1
 /product="source of immunodominant MHC-associated
 peptides"
 /protein_id="AAL71884.1"
 /db_xref="GI:19879589"
 /translation="MAEPAPESKHKSSLSNPSGLMGNLSRHGHGPGCAQAHKA

AGGAAPKAPAGLGGCLSPACWQSLISFTILFLAWLAGFSRSLFAVIRESIHEF
 DPWNYSTHLLASHGFEPLNWFDERAWYPLGRIVGCTVPLGLTAGLJHMLNLT
 NITVHSDVCVFLAATPFGSLTISITELLRELWNGALLAACFAIATVPGIISRSVAG
 SFDNEGIAIFALQFTYILWVKSVTGSFWMCCCLSYFYFVMSAWGFFVILNPLH
 VFLLLRYSKRKYIAYSTFYVGLILSMQIPFVGFQPIRTSEMAAGAFVALLQAY
 AFLGLDRLTQBFOTLFFGLVLAAGVFLSYLYTGYIAEWSGRFYSLMDTGY
 AKIHPIIASVSHQPTTWSPFFDLHLIVCTFFPAGLWFCIKINDEKRVALLAISA
 VYFAGVVRMLTLTPVCMLSAIAFSNVFHYLGDMMKRNPPVEDSDSDKRNQO
 NLXKACKVRKHATEOEKTEEGGPNKSIIVTMLMLMLMFAVCTVWNTNAYSSPS
 VLASVNHGDTNMLDDEPBAVFWLQNTDEHARVMSWDYGYQIAGNARVTLVDNN
 TANNSHALVCKMSSNEHAYKIMETLVDYVLIYFSGVIGYSGDDINKELMWURIA
 EGEHPKIDRESDIYTPGGEFRVDKAGSPILLNLMKMSIYRFGSQDLQFPIPGFDR
 TRNAIEGNKIDKFRHAEAFTESEHMLVRIYKVKAPDNRETLDHKPRVTNIPFKQKYL
 KKTTRKRGYIKNKLFFKKGKISKVTV"

ORIGIN

Query Match 100.0%; Score 2481; DB 9; Length 2481;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGGAGCCCTCGGCGCCCGGAGAGCAAGCAAACTGCTCCTCAACTGCTCCCGTGG 60
 DB 1 ATGCGGAGCCCTCGGCGCCCGGAGAGCAAGCAAACTGCTCCTCAACTGCTCCCGTGG 60
 QY 61 AGTGGCCTCATGCGCCTGGGAAACAGCCGCGCACCGCCGCGGCGCCCGGCGCCAGTGC 120
 DB 61 AGTGGCCTCATGCGCCTGGGAAACAGCCGCGCACCGCCGCGGCGCCCGGCGCCAGTGC 120
 QY 121 GCGCAAGAGCG 180
 DB 121 GCGCAAGAGCG 180
 QY 181 GGGCTGTGCGAGCGCGTGGGTGGCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
 DB 181 GGGCTGTGCGAGCGCGTGGGTGGCGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
 QY 241 TGGCTTGGCGGCTTCACTGCGCGCTTTCGCGCTTTCGCGCTTTCGCGCTTTCGCGCTTTC 300
 DB 241 TGGCTTGGCGGCTTCACTGCGCGCTTTCGCGCTTTCGCGCTTTCGCGCTTTCGCGCTTTC 300
 QY 301 GAGTTCGACCGCGTGGTTAACTATAGATCAACACATCATCTTGCATCTCATGGTCTTAT 360
 DB 301 GAGTTCGACCGCGTGGTTAACTATAGATCAACACATCATCTTGCATCTCATGGTCTTAT 360
 QY 361 GAATTTTAAATTTGGTTGATGAAGAGCATGTATCCACTAGGAGATAGTAGTGGT 420
 DB 361 GAATTTTAAATTTGGTTGATGAAGAGCATGTATCCACTAGGAGATAGTAGTGGT 420
 QY 421 ACTGTTTACCCAGGTTGATGATAAACCCTGGCTTATTCATTGGATTTTAAATACATTG 480
 DB 421 ACTGTTTACCCAGGTTGATGATAAACCCTGGCTTATTCATTGGATTTTAAATACATTG 480
 QY 481 ACATTAATCTGTTACATGAAGAGCATGTATGCTTCTCTGCGACCACTTTTACCGGCTT 540
 DB 481 ACATTAATCTGTTACATGAAGAGCATGTATGCTTCTCTGCGACCACTTTTACCGGCTT 540
 QY 541 ACATCTATATCTACTTTCTCTGCTTACAGAGAACTTTTGGAAACCAAGGAGGAGACTTTTA 600
 DB 541 ACATCTATATCTACTTTCTCTGCTTACAGAGAACTTTTGGAAACCAAGGAGGAGACTTTTA 600
 QY 601 GCTGCTGTTTTTATTTGCTATTGTACAGGCTCATATCTCGGTGAGTAGTGGTCCCTT 660
 DB 601 GCTGCTGTTTTTATTTGCTATTGTACAGGCTCATATCTCGGTGAGTAGTGGTCCCTT 660
 QY 661 GATAATGAAGCATGCTATTTTGTGACTTCAGTTCACATCTATTTATGGTAAATCT 720
 DB 661 GATAATGAAGCATGCTATTTTGTGACTTCAGTTCACATCTATTTATGGTAAATCT 720
 QY 721 GTAAATACTGGGTGAGTTTTTGTGACAAATGTGCTGCTTATCTTATTTCTATATGTC 780
 DB 721 GTAAATACTGGGTGAGTTTTTGTGACAAATGTGCTGCTTATCTTATTTCTATATGTC 780
 QY 781 TCTGCTGGGTGGTTATGATTAATTAATCAATCTTATTCACCTGATGATTTGTTG 840

ORIGIN

Query Match 100.0%; Score 2481; DB 6; Length 2481;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGAGCCCTCGGCGCGGAGCAAGCAAGTCTCCCTCAACTCGTCCCGTGG 60
DB 1 ATGGCGAGCCCTCGGCGCGGAGCAAGCAAGTCTCCCTCAACTCGTCCCGTGG 60

QY 61 AGTGGCTCATGGCCCTGGGAAACAGCCGCGCAACAGCGGCGCGGCGCGGCTGC 120
DB 61 AGTGGCTCATGGCCCTGGGAAACAGCCGCGCAACAGCGGCGCGGCGCGGCTGC 120

QY 121 GCGCAAGAGCGGCGGCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 121 GCGCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180

QY 181 GGGCTGTCGAGCGCGGCTGGGTGGCGAGTCTCTCTCTTCAACATCTCTTCTGGCC 240
DB 181 GGGCTGTCGAGCGCGGCTGGGTGGCGAGTCTCTCTCTTCAACATCTCTTCTGGCC 240

QY 241 TGGCTTGGCGGCTTCACTCGGCTCTTCCGCGTCATCGCTTCCGAAAGCATATCCAC 300
DB 241 TGGCTTGGCGGCTTCACTCGGCTCTTCCGCGTCATCGCTTCCGAAAGCATATCCAC 300

QY 301 GAGTTCGAGCGGCTGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGGTTCTAT 360
DB 301 GAGTTCGAGCGGCTGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGGTTCTAT 360

QY 361 GAATTTTAAATGGTTTGTATGAAGAGCATGTATCCATAGGAAAGATAGTAGGTGGT 420
DB 361 GAATTTTAAATGGTTTGTATGAAGAGCATGTATCCATAGGAAAGATAGTAGGTGGT 420

QY 421 ACTGTTTACCCAGGTTGATGATAACCGCTGGCTTATTCAITGGATTTTAAATACATTG 480
DB 421 ACTGTTTACCCAGGTTGATGATAACCGCTGGCTTATTCAITGGATTTTAAATACATTG 480

QY 481 AACATATCTTCACTATGAAGAGCATGTATGTTGCTTGGCACCACATTTTACGGGCTTT 540
DB 481 AACATATCTTCACTATGAAGAGCATGTATGTTGCTTGGCACCACATTTTACGGGCTTT 540

QY 541 ACATCTATATCTCTTCTTCTTCAAGAGAACTTTGGAACCAAGGACAGACTTTTA 600
DB 541 ACATCTATATCTCTTCTTCTTCAAGAGAACTTTGGAACCAAGGACAGACTTTTA 600

QY 601 GCTGCTGTTTATGCTATGTACAGGCTATATCTCGCTCAGTACGCTGATCTCTTT 660
DB 601 GCTGCTGTTTATGCTATGTACAGGCTATATCTCGCTCAGTACGCTGATCTCTTT 660

QY 661 GATAATGAAGCATTTGCTATTTTGGCACTTCAGTTCACATATATTTATGGGTAAATCT 720
DB 661 GATAATGAAGCATTTGCTATTTTGGCACTTCAGTTCACATATATTTATGGGTAAATCT 720

QY 721 GTAAAGCTGGTTCAGTTTGGCAATGTGCTGCTTATCTTCTATATGCTC 780
DB 721 GTAAAGCTGGTTCAGTTTGGCAATGTGCTGCTTATCTTCTATATGCTC 780

QY 781 TCTGCTTGGGTTGTTATGTTATCATCAATCTTATTTCCACTGCATGTTTGTGTTG 840
DB 781 TCTGCTTGGGTTGTTATGTTATCATCAATCTTATTTCCACTGCATGTTTGTGTTG 840

QY 841 TTACTGTAGCAGATACAGCAAGAGTCTACATAGCATATAGCATCTTCTACATTTG 900
DB 841 TTACTGTAGCAGATACAGCAAGAGTCTACATAGCATATAGCATCTTCTACATTTG 900

QY 901 GGTTTAATATTAATCAATGAGATACCTTTTGTGGATTTCCAGCAATCAAGAGTAA 960
DB 901 GGTTTAATATTAATCAATGAGATACCTTTTGTGGATTTCCAGCAATCAAGAGTAA 960

QY 961 CACATGGCAGTGCAGGTGCTTGTGATTCGCAAGCTTATGCTTCTTGTGAGTATCTG 1020
DB 961 CACATGGCAGTGCAGGTGCTTGTGATTCGCAAGCTTATGCTTCTTGTGAGTATCTG 1020

QY 1021 AGAGACCGATTAAACAAAACAGGTTCCAGACCCCTTTCTTTTGGGTGATCACTAGCT 1080
DB 1021 AGAGACCGATTAAACAAAACAGGTTCCAGACCCCTTTCTTTTGGGTGATCACTAGCT 1080

QY 1081 GCAGGTGCTGTGTTTCTTTAGTGTCTATTTTGAATTTATACAGTTACATTTGACCATGG 1140
DB 1081 GCAGGTGCTGTGTTTCTTTAGTGTCTATTTTGAATTTATACAGTTACATTTGACCATGG 1140

QY 1141 AGTGGCAGGTTTATTTATTTGGGATCTGGGTATCAAAAATACACATTCCTCAATTTAT 1200
DB 1141 AGTGGCAGGTTTATTTATTTGGGATCTGGGTATCAAAAATACACATTCCTCAATTTAT 1200

QY 1201 GCATCAGTGTCTGAGCATCAACCTACGACTTTGGGTGCTTTCTTTTGTATCTACATATT 1260
DB 1201 GCATCAGTGTCTGAGCATCAACCTACGACTTTGGGTGCTTTCTTTTGTATCTACATATT 1260

QY 1261 CTTGTATGTACCTTCCAGCAGGCTTTGGTCTGTCATCAAAAATATCAACGATGAAGA 1320
DB 1261 CTTGTATGTACCTTCCAGCAGGCTTTGGTCTGTCATCAAAAATATCAACGATGAAGA 1320

QY 1321 GTATTTGTTGCTCTATATGCAATCAGTGTCTCTACTTTGCTGGAGTATGTTGCGACTG 1380
DB 1321 GTATTTGTTGCTCTATATGCAATCAGTGTCTCTACTTTGCTGGAGTATGTTGCGACTG 1380

QY 1381 ATGTTGACTTTGACTTCCAGTGTGTATGCTGTCTGCAATTTGCTTTTCAAAATGTTTTT 1440
DB 1381 ATGTTGACTTTGACTTCCAGTGTGTATGCTGTCTGCAATTTGCTTTTCAAAATGTTTTT 1440

QY 1441 GAGCACTATTTGGGGATGACATGAAAAGGAAAAATCCACTGTGGAGGACAGCTGAT 1500
DB 1441 GAGCACTATTTGGGGATGACATGAAAAGGAAAAATCCACTGTGGAGGACAGCTGAT 1500

QY 1501 GAGGATGACAAAAGAAAACCAAGGAAATTTGTATGATAAGGACAGTAAAGTGAGGAAACAT 1560
DB 1501 GAGGATGACAAAAGAAAACCAAGGAAATTTGTATGATAAGGACAGTAAAGTGAGGAAACAT 1560

QY 1561 GCACCTGACAGAAAACCAAGGATTTAGGCTTATTAATAAGGATGTCACC 1620
DB 1561 GCACCTGACAGAAAACCAAGGATTTAGGCTTATTAATAAGGATGTCACC 1620

QY 1621 ATGTTGATGCTGATCTATTTGATGATGTTTGTCTGCTTCCACTGTACCTGGGTCAACAACAT 1680
DB 1621 ATGTTGATGCTGATCTATTTGATGATGTTTGTCTGCTTCCACTGTACCTGGGTCAACAACAT 1680

QY 1681 GCTACTCTAGTCCAGGTGATGCTGCTCATACATCATGATGACGACAGGATATC 1740
DB 1681 GCTACTCTAGTCCAGGTGATGCTGCTCATACATCATGATGACGACAGGATATC 1740

QY 1741 TTAGATGATTTTGAAGAAGCTTACTTTTGGCTAAGGCAAAAATACAGATGAACATGACCGA 1800
DB 1741 TTAGATGATTTTGAAGAAGCTTACTTTTGGCTAAGGCAAAAATACAGATGAACATGACCGA 1800

QY 1801 GTAAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGAATAGAACTAGCTTTC 1860
DB 1801 GTAAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGAATAGAACTAGCTTTC 1860

QY 1861 GTGGATTAATAACCTGGAATAACAGCCACATAGCACTGTGGGAAAGCTATGCTTCT 1920
DB 1861 GTGGATTAATAACCTGGAATAACAGCCACATAGCACTGTGGGAAAGCTATGCTTCT 1920

QY 1921 AATGAAACAGCAGCTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTTATT 1980
DB 1921 AATGAAACAGCAGCTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTTATT 1980

QY 1981 TTTGAGGGGTTATTGGCTATTTCTGGTGTATGATTAACAAATTTCTCTGGATGGTTAGG 2040
DB 1981 TTTGAGGGGTTATTGGCTATTTCTGGTGTATGATTAACAAATTTCTCTGGATGGTTAGG 2040

QY 2041 ATAGCTGGAAGGAGAAACATCCCAAGACATTCGGAAGAGTACTATTTTATCCCAACAGGA 2100
DB 2041 ATAGCTGGAAGGAGAAACATCCCAAGACATTCGGAAGAGTACTATTTTATCCCAACAGGA 2100

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 02:41:23 ; Search time: 16661.4 Seconds
(without alignments)
7041.770 Million cell updates/sec

Title: US-10-028-384-1

Perfect score: 2481

Sequence: 1 atggcgagcctcgccccc.....tatctaagaagactgtttaa 2481

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2481	100.0	2481	6 AX799082	AX799082 Sequence
2	2481	100.0	2481	9 AY074880	AY074880 Homo sapi
3	1276	51.4	1664	6 CQ782554	CQ782554 Sequence
4	1276	51.4	1664	6 BD127193	BD127193 Primer fo
5	1276	51.4	1664	9 AK074587	AK074587 Homo sapi
6	972	39.2	2546	6 AX099510	AX099510 Sequence
7	972	39.2	2546	6 BD063986	BD063986 Secreted
8	954	38.5	2510	6 AX136135	AX136135 Sequence
9	954	38.5	2510	6 BD123520	BD123520 Secretory
10	954	38.5	2510	9 AK075380	AK075380 Homo sapi
11	861	34.7	2284	6 AX882932	AX882932 Sequence
12	861	34.7	2284	6 BD160013	BD160013 Primer fo
13	861	34.7	2284	9 AK027789	AK027789 Homo sapi
14	828	33.4	2503	6 AX017997	AX017997 Sequence
15	828	33.4	2503	6 BD137303	BD137303 Human nuc
16	739	29.8	957	9 BC015880	BC015880 Homo sapi
17	521	21.0	764	6 AX136480	AX136480 Sequence
18	521	21.0	764	6 BD123720	BD123720 Secretory
19	507	20.4	507	6 AX886102	AX886102 Sequence

20	507	20.4	507	6 BD025712	BD025712 Sequence
21	506	20.4	787	6 AX869456	AX869456 Sequence
22	506	20.4	787	6 BD149518	BD149518 Primer fo
23	502	20.2	1040	6 CQ723424	CQ723424 Sequence
24	430	17.3	433	6 AX340470	AX340470 Sequence
25	402	16.2	558	6 AX079689	AX079689 Sequence
26	351	14.1	616	6 BD210442	BD210442 Human gen
27	322	13.0	349	6 BD071886	BD071886 Secreted
28	316	12.7	170691	9 AC104643	AC104643 Homo sapi
29	250	10.1	250	6 AX885575	AX885575 Sequence
30	250	10.1	250	6 BD025185	BD025185 Sequence
31	236	9.5	245	6 AX905227	AX905227 Sequence
32	236	9.5	245	6 BD040760	BD040760 Sequence
33	236	9.5	247	6 BD075907	BD075907 5' EST of
34	218	8.8	503	6 CQ779964	CQ779964 Sequence
35	218	8.8	503	6 CQ781897	CQ781897 Sequence
36	218	8.8	503	6 BD124673	BD124673 Primer fo
37	218	8.8	503	6 BD126606	BD126606 Primer fo
38	218	8.8	177253	9 AC092024	AC092024 Homo sapi
39	188	7.6	388	6 AX071612	AX071612 Sequence
40	184	7.4	737	6 BD020608	BD020608 Novel gen
41	184	7.4	737	6 BD100546	BD100546 Novel gen
42	179	7.2	474	6 AX079710	AX079710 Sequence
43	170	6.9	835	9 HSA324805	AJ324805 Homo sapi
44	124	5.0	355	6 AX340785	AX340785 Sequence
45	114	4.6	567	6 CQ780783	CQ780783 Sequence

ALIGNMENTS

RESULT 1	AX799082	2481 bp	mRNA	linear	PAT 08-OCT-2003
LOCUS	AX799082	Sequence 1 from Patent WO03054008.			
DEFINITION	AX799082				
ACCESSION	AX799082.1	GI:37605057			
VERSION					
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	Perreault, C. and McBride, K.				
AUTHORS	Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy				
TITLE	Patent: WO 03054008-A 1 03-JUL-2003;				
JOURNAL	Comptatigene Inc. (CA)				
FEATURES	Location/Qualifiers				
source	1..2481				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
CDS	1..2481				
	/codon_start=1				
	/product="SIMP"				
	/protein_id="CAE48447.1"				
	/db_xref="GI:37605058"				
	/translation="MAEPPAPESKHKSLNSPWSGLMALGNSRHCHGHPGACCAHKA AGAAPKPPAPAGLSGLSQPAGWQSLISFTILFLAWLAGFSRLFAVIRPESITHEF DPWFYRSTHLLASHGCFEFLNWFDPRAWVPLGRIVGTVYGLMITAGLHWLNTL NITHIDVCVFLAPTFSGLTISITELLRELNQAGLLAACFAIVGVTSRVAG SFDEGTAIFALQFTIYLWYKSVKTSVFWIMCCCLSYFIRVSEHNAAGVALLQAY FVLLLMORVSKRYIAYSTFYIVGLILNQIPFVGFQPIRSEHNAAGVALLQAY AFLQYLRDLTKQEFQTLFELGSLAAGAVLSVLYLTGTICINDEKRVFVALYISA AKIHPIIASVSEHQPTWTFPFDFLHVCTFFPAGLWFCIKINDEKRVFVALYISA VYFAGVMVRLMLTLPVVCMLSAFNSVFEHLGDMKRENPVPEVSDSDKRNQ NLVDKAGKVRKHATQEQTEEGLEGPKISIVTLMMLMFMVACHTWTNAYSQPS VVLASVHNDGTRNILDREAFVFWLQNTDEHARVMSWDYGVYAGNAGMTLVLDNN TWNSHVALVKAMSNSETAAVKIMETLDDVVLVIFGVGIVGSDGDKINKLFWVIA EGEHPKDIRSDYFTQGEPRVDKAGSPILLNCLMYKMSYVFGQMQLDFTPGPDR TRNEIGNKDKIKHLEEAFTSEHNLVRYKVKAPDNRETLDHKEKRVNIFPKQKLS KKTTRKRKRGYIKNKLVEFKKKLSKKTIV"				

~~Dupli~~

Blank Page

Db	458	GATATCAACAAATTCCTGGAATGGTTAGGATAGCTGAAGGAGAACATCCCAAGACATT	517
Qy	2071	CGGGAAGTGACTATTTACCCACAGGAGAAATTCGGTGTAGACAAAGCAGGATCCCT	2130
Db	518	CGGGAAGTGACTATTTACCCACAGGAGAAATTCGGTGTAGACAAAGCAGGATCCCT	577
Qy	2131	ACTTTGTTGAATTCGCTTATGTATAAAATGTCATACATGATTTGGAGAAATGCAGCTG	2190
Db	578	ACTTTGKTGANTGCTTATGTATAAAATGTCATACATGATTTGGAGAAATGCAGCTG	637
Qy	2191	GATTTTCGTACACCCCGAGGTTTGGACCGAACACGTAATGCTGAGATTGGAATTAAGGAC	2250
Db	638	GATTTTCGTACACCCCGAGGTTTGGACCGAACACGTAATGCTGAGATTGGAATTAAGGAC	697
Qy	2251	ATTAATTCAAACATTTGGAGAGCCCTTTACATCAGAACACTGGCTTGTAGGATATAT	2310
Db	698	ATTAATTCAAACATTTGGAGAGCCCTTTACATCAGAACACTGGCTTGTAGGATATAT	757
Qy	2311	AAAGTAAAGCACCTGTATACAGGGAGACATTTAGATCACAAACCTCGAGTCCCAACATT	2370
Db	758	AAAGTAAAGCACCTGTATACAGGGAGACATTTAGATCACAAACCTCGAGTCCCAACATT	817
Qy	2371	TTCCCAAAACAGAAAGTATTTGTCAAAGAGACTACCAAAGGAGCGTGGCTACATTAA	2430
Db	818	TTCCCAAAACAGAAAGTATTTGTCAAAGAGACTACCAAAGGAGCGTGGCTACATTAA	877
Qy	2431	AATAAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAGAGACTGTTTTAA	2481
Db	878	AATAAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAGAGACTGTTTTAA	928

Search completed: December 15, 2004, 15:46:57

Job time : 1672 secs

```

Db 543 TCATCTACAGATTGGGAAATGAGTGGATTTTCGTACCCAGGTTTGACCGA 602
QY 2221 ACACGTAATGCTGAGATTGGAATAGGACATTAAATTTCAACATTGGAAGAGCTTT 2280
Db 603 ACACGTAATGCTGAGATTGGAATAGGACATTAAATTTCAACATTGGAAGAGCTTT 662
QY 2281 ACATCAGAACACTGGCTGTTAGGATATATAAGTAAAGACACCTGATACAGGGAGCA 2340
Db 663 ACATCAGAACACTGGCTGTTAGGATATATAAGTAAAGACACCTGATACAGGGAGCA 722
QY 2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCAAAACAGAGTATTGTCAAGAAG 2400
Db 723 TTAGATCAAAACCTCGAGTCACCAACATTTTCCAAAACAGAGTATTGTCAAGAAG 782
QY 2401 ACTACCAAAGGAGGCTGGCTACATTAAATAAGCTGTTTAAAGAAAGGCAAGAA 2460
Db 783 ACTACCAAAGGAGGCTGGCTACATTAAATAAGCTGTTTAAAGAAAGGCAAGAA 842
QY 2461 ATAATCAAGAGCTGTTTAA 2481
Db 843 ATAATCAAGAGCTGTTTAA 863

RESULT 15
AAD08315
ID AAD08315 standard; cDNA; 2547 BP.
AC AAD08315;
DT 08-AUG-2001 (first entry)
DE Human secreted protein-encoding gene 7 cDNA clone HDTLR06, SEQ ID NO: 43.
KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vunerary; binding partner identification;
KW gene therapy; ss.
XX Homo sapiens.
XX OS
XX FH
XX Key Location/Qualifiers
XX CDS 68..928
XX /*tag= a
XX /product= "Human secreted protein precursor"
XX /trans_except= (pos:584..586, aa:xa)
XX /note= "Xaa equals any of the naturally occurring L-amino
XX acids"
XX sig_peptide 68..130
XX /*tag= b
XX mat_peptide 131..925
XX /*tag= c
XX /product= "Mature human secreted protein"
XX
XX WO200136440-A1.
XX PN
XX PD
XX 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-US031282.
XX PF
XX 19-NOV-1999; 99US-0166414P.
XX PR
XX 21-JUL-2000; 2000US-0219655P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX Ruben SM, Komatsoulis GA, Birse CE, Moore PA;
XX PI
XX

```

```

DR WPI; 2001-343795/36.
DR P-PSDB; AAE03850.
XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.
XX Claim 1; Page 462-463; 553pp; English.
XX
XX AAD08283-AAD08355 represent cDNAs corresponding to 23 human secreted
XX protein genes, and AAD03818-AAD03870 represent the proteins they encode.
XX AAD03871-AAD03896 represent human secreted protein fragments or variants.
XX The secreted proteins and their genes are useful for preventing, treating
XX or ameliorating medical conditions, e.g., by protein or gene therapy.
XX Pathological conditions can be diagnosed by determining the amount of the
XX new protein in a sample or by determining the presence of mutations in
XX the new genes. Specific uses are described for each of the 23 genes,
XX based on the tissues in which they are most highly expressed, and include
XX developing products for the diagnosis or treatment of proliferative
XX disorders, cancer, tumours, foetal and developmental abnormalities,
XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
XX angiogenic disorders, kidney disorders, gastrointestinal disorders,
XX pregnancy-related disorders, endocrine disorders, and infections. The
XX proteins can also be used to aid wound healing and epithelial cell
XX proliferation, to prevent skin aging due to sunburn, to maintain organs
XX before transplantation, for supporting cell culture of primary tissues,
XX to regenerate tissues, to identify their cognate ligands or binding
XX partners, and in chemotaxis, and can be used as a food additive or
XX preservative to modify storage properties. Antibodies specific for a
XX protein of the invention can be used in alleviating symptoms associated
XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,
XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
XX present sequence represents a human secreted protein-encoding cDNA of the
XX invention
XX
XX SQ Sequence 2547 BP; 833 A; 422 C; 492 G; 798 T; 0 U; 2 Other;
XX
XX Query Match 33.9%; Score 840; DB 4; Length 2547;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 890; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1591 TTAGGCCCTAATAATAAAAGCATTGTGCACCATGTTGTGCTGATGCTATTGATGATGTTT 1650
Db 38 TTAGGCCCTAATAATAAAAGCATTGTGCACCATGTTGTGCTGATGCTATTGATGATGTTT 97
QY 1651 GCTGTCCACTGTACTGGGTCACAGCAATGCTTACTCTAGTCCAGGTAGTCTCGCC 1710
Db 98 GCTGTCCACTGTACTGGGTCACAGCAATGCTTACTCTAGTCCAGGTAGTCTCGCC 157
QY 1711 TCATACAATCATGATGGCCACCAGGAATATCTTAGATGATTTTAGAGAAGCTTACTTTGG 1770
Db 158 TCATACAATCATGATGGCCACCAGGAATATCTTAGATGATTTTAGAGAAGCTTACTTTGG 217
QY 1771 CTAAGGCAAAATACAGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1830
Db 218 CTAAGGCAAAATACAGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 277
QY 1831 ATAGCTGGAATGGCTAATAGAACTACGTTGGTGGATAATAACACCTGGAATAACAGCCAC 1890
Db 278 ATAGCTGGAATGGCTAATAGAACTACGTTGGTGGATAATAACACCTGGAATAACAGCCAC 337
QY 1891 ATAGCACTGGTGGGAAAGCTATGCTTCTAATGAACAGCAGCCTATAAATCATGAGG 1950
Db 338 ATAGCACTGGTGGGAAAGCTATGCTTCTAATGAACAGCAGCCTATAAATCATGAGG 397
QY 1951 ACTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2010
Db 398 ACTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 457
QY 2011 GRTATCAACAAATTTCTCTGGATGTTAGGTAGTCTGAAGGAGACATCCCAAGACATT 2070

```


QY 1810 TGGTGGGATTATCGCTATCAGATAGCTGGATGGCTTAATAGAACTACCTTGGTGGATAAT 1869
 Db 301 TGGTGGGATTATGGCTATCAGATAGCTGGATGGCTTAATAGAACTACCTTGGTGGATAAT 360
 QY 1870 AACACCTGGGAATACAGCCACATAGCACTGGTGGGAAAGCTATGCTCTTCTAATGAACA 1929
 Db 361 AACACCTGGGAATACAGCCACATAGCACTGGTGGGAAAGCTATGCTCTTCTAATGAACA 420
 QY 1930 GCACCTATTAATATCAGAGCTAGATAGATATGTTTGGTATTTTGGAGGG 1989
 Db 421 GCACCTATTAATATCAGAGCTAGATAGATATGTTTGGTATTTTGGAGGG 480
 QY 1990 GTTATTGGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTATAGATAGCTGAA 2049
 Db 481 GTTATTGGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTATAGATAGCTGAA 540
 QY 2050 GGAGACATCCCAAGACATTCGGGAAGTGAATTTTACCCACAGGGAGNATCCGT 2109
 Db 541 GGAGACATCCCAAGACATTCGGGAAGTGAATTTTACCCACAGGGAGNATCCGT 600
 QY 2110 GTACACAAAGCAGATCCCTACTTTTGTGAATTCCTTATGATATAAATGTCATCTAC 2169
 Db 601 GTACACAAAGCAGATCCCTACTTTTGTGAATTCCTTATGATATAAATGTCATCTAC 660
 QY 2170 AGATTGGGAATTCGAGCTGGATTTTCGTACACCCCGAGTTTTCGACGACAGCTAAT 2229
 Db 661 AGATTGGGAATTCGAGCTGGATTTTCGTACACCCCGAGTTTTCGACGACAGCTAAT 720
 QY 2230 GCTGAGATTGGAATTAAGACATTAATTAATCAACATTTTGGAGAGCCTTTACATCAGAA 2289
 Db 721 GCTGAGATTGGAATTAAGACATTAATTAATCAACATTTTGGAGAGCCTTTACATCAGAA 780
 QY 2290 CACTGGCTTGTAGATATATAAGTAAGCAACCTGTAAACAGGAGACATATAGATCAC 2349
 Db 781 CACTGGCTTGTAGATATATAAGTAAGCAACCTGTAAACAGGAGACATATAGATCAC 840
 QY 2350 AAACCTCGAGTACCAACATTTTCCCAAAACAGAGATATTTTCAAGAGACATACCAA 2409
 Db 841 AAACCTCGAGTACCAACATTTTCCCAAAACAGAGATATTTTCAAGAGACATACCAA 900
 QY 2410 AGGAAGCTGGCTACATTAATAAATAGCTGGTTTTCAGAAAGGCAAGAAATATCTAAG 2469
 Db 901 AGGAAGCTGGCTACATTAATAAATAGCTGGTTTTCAGAAAGGCAAGAAATATCTAAG 960
 QY 2470 AAGACTGTTAA 2481
 Db 961 AAGACTGTTAA 972

RESULT 12

AAF93772
 ID AAF93772 standard; cDNA; 2510 BP.
 XX
 AC AAF93772;
 XX
 DT 23-MAY-2001 (first entry)
 XX
 DE Human cDNA encoding a membrane or secretory protein clone PSE0070.
 XX
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1067182-A2.
 XX
 PD 10-JAN-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114090.
 XX
 PR 08-JUL-1999; 99JP-00194179.
 PR 11-JAN-2000; 2000JP-00118775.
 PR 02-MAY-2000; 2000JP-00183766.
 PR

XX (HELI-) HELIX RES INST.
 PA Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 PI WPI; 2001-093989/11.
 XX P-PSDB; AAB88345.
 DR
 DR
 XX
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development.
 XX
 PS Claim 1; SEQ ID NO 57; 609pp + Sequence Listing; English.
 XX
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by AAB88317
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
 CC invention. The invention also includes methods for the production of
 CC antibodies directed against the proteins, and cDNA sequences, which can
 CC be used in vaccines. The polynucleotide sequences can be used in gene
 CC therapy. The polynucleotide sequences and the proteins they encode may be
 CC used in the prevention, treatment and diagnosis of diseases associated
 CC with inappropriate secretory protein/membrane protein expression. The
 CC nucleic acids and complementary sequences may also be used as DNA probes
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
 CC and quantitate the presence of similar nucleic acid sequences in samples.
 CC They may also be used to study the expression and function of secretory
 CC proteins/membrane polypeptides and their role in metabolism. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC against them and in assays to identify modulators (agonists and
 CC antagonists) of expression and activity. The antibodies and antagonists
 CC may also be used as therapeutic agents to down regulate expression and
 CC activity. The antibodies may also be used as diagnostic agents for
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme
 CC linked immunosorbent assay (ELISA)). Examples of diseases which may be
 CC treated include rheumatoid arthritis and diabetes
 XX
 SQ Sequence 2510 BP; 802 A; 412 C; 489 G; 807 T; 0 U; 0 Other;
 Query Match 38.5%; Score 954; DB 5; Length 2510;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1528 TTGTTATGATTAAGGAGGTAAGTGAAGAAACATGCACTGCAAGAAAGAACTGAAGAG 1587
 Db 1 TTGTTATGATTAAGGAGGTAAGTGAAGAAACATGCACTGCAAGAAAGAAAGAACTGAAGAG 60
 QY 1588 GGATTAGGCCCTTAATATAAAAGCATTGTCCACATGTTGATGCTGATGCTATTTGATGATG 1647
 Db 61 GGATTAGGCCCTTAATATAAAAGCATTGTCCACATGTTGATGCTGATGCTATTTGATGATG 120
 QY 1648 TTGCTGTCCACTGTACTGGCTCAGAGCAATGCCCTACTCTAGTCCAGTGTAGTCCCTG 1707
 Db 121 TTGCTGTCCACTGTACTGGCTCAGAGCAATGCCCTACTCTAGTCCAGTGTAGTCCCTG 180
 QY 1708 GCCTCATACATCATGATGCGCACCAGCAATATCTTAGATGATTTTAGAGAAGCTTACTTT 1767
 Db 181 GCCTCATACATCATGATGCGCACCAGCAATATCTTAGATGATTTTAGAGAAGCTTACTTT 240
 QY 1768 TGGCTAAGGCAAAATACAGATGACATGACATGACATGACATGACATGACATGACATGAC 1827
 Db 241 TGGCTAAGGCAAAATACAGATGACATGACATGACATGACATGACATGACATGACATGAC 300
 QY 1828 CAGATAGCTGGAATGGCTTAATAGAACTACCGTTGGTGAATAATTAACACCTGGAATAACAG 1887
 Db 301 CAGATAGCTGGAATGGCTTAATAGAACTACCGTTGGTGAATAATTAACACCTGGAATAACAG 360
 QY 1888 CACATAGCACTGGTGGGAAAGCTATGCTCTTCTAATGAACACAGCAGCTTAATATCATG 1947
 Db 361 CACATAGCACTGGTGGGAAAGCTATGCTCTTCTAATGAACACAGCAGCTTAATATCATG 420
 QY 1948 AGGACTCTAGATGATGATATGTTTGGTATTTTGGAGGGTATTTGGCTATTTCTGTT 2007

1570	QY	CAGGAAAAAACTGAAGAGGATAGGCCCTAATATAAAAGCAATTGTCCACCATTTGATG	1629
61	Db	CAGGAAAAAACTGAAGAGGATAGGCCCTAATATAAAAGCAATTGTCCACCATTTGATG	120
1630	QY	CTGATGCTATTGATGATGTTTGCTGTCCACTGTACTCTGGGTCAACAGCAATGCCTACTCT	1689
121	Db	CTGATGCTATTGATGATGTTTGCTGTCCACTGTACTCTGGGTCAACAGCAATGCCTACTCT	180
1690	QY	AGTCCAGGTGAGTCTCTGGCCCTCATACAACTCATGATGGCACCGGATATCTTTAGATGAT	1749
181	Db	AGTCCAGGTGAGTCTCTGGCCCTCATACAACTCATGATGGCACCGGATATCTTTAGATGAT	240
1750	QY	TTTAGAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGCACGAGTAATGCTCT	1809
241	Db	TTTAGAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGCACGAGTAATGCTCT	300
1810	QY	TGTTGGGATATGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGGTGGATAAT	1869
301	Db	TGTTGGGNTATGCTATCAGATAGCTGGNATGGCTAATAGAACTACGTTGGTGGATAAT	360
1870	QY	AACACTTGGAAATAACAGCCACATAGACATGCTGGTGGGAAAGCTATGTCTTTATGAAACA	1929
361	Db	AACACTTGGAAATAACAGCCACATAGACATGCTGGTGGGAAAGCTATGTCTTTATGAAACA	420
1930	QY	GCAGCTATAAAATCATGAGGACTCTAGATGTAGATTATGTTTGGTTATTTTGGAGGG	1989
421	Db	GCAGCTATAAAATCATGAGGACTCTAGATGTAGATTATGTTTGGTTATTTTGGAGGG	480
1990	QY	GTTATTGGCTATTCTGGTGATATACAAATTTCTCTGGATGGTTAGGATAGCTGAA	2049
481	Db	GTTATTGGCTATTCTGGTGATATACAAATTTCTCTGGATGGTTAGGATAGCTGAA	540
2050	QY	GGAGAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCACAGGGAGAAATCCGT	2109
541	Db	GGAGAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCACAGGGAGAAATCCGT	600
2110	QY	GTAGACAAGCAGGATCCCTACTTTTGGATTTGCTATGTATAAAATGCTCATCTAC	2169
601	Db	GTAGACAAGCAGGATCCCTACTTTTGGATTTGCTATGTATAAAATGCTCATCTAC	660
2170	QY	AGATTTGAGAAATGCAAGCTGGATTTTGGTACACCCCGAGTTTGGACCGAACATGPAAT	2229
661	Db	AGATTTGAGAAATGCAAGCTGGATTTTGGTACACCCCGAGTTTGGACCGAACATGPAAT	720
2230	QY	GCTGAGATTGGAAATAAGGACATTAATTAATCAAACTTTGGAGAGGCTTTACATCAGAA	2289
721	Db	GCTGAGATTGGAAATAAGGACATTAATTAATCAAACTTTGGAGAGGCTTTACATCAGAA	780
2290	QY	CAGTGGCTTTGTTAGGATATATAAAGTAAAGCACTGTATTAACAGGGAGACATTAGATCAC	2349
781	Db	CAGTGGCTTTGTTAGGATATATAAAGTAAAGCACTGTATTAACAGGGAGACATTAGATCAC	840
2350	QY	AAAGCTCGAGTCACCAACATTTTCCCAAAACAGAGTATTTGTCAAAAGAGACTTACCAA	2409
841	Db	AAAGCTCGAGTCACCAACATTTTCCCAAAACAGAGTATTTGTCAAAAGAGACTTACCAA	900
2410	QY	AGGAAGCTGCTACATTAATAATAGCTGGTTTTTAAAGAGCCAGAAATATCTAAG	2469
901	Db	AGGAAGCTGCTGCTACATTAATAATAGCTGGTTTTTAAAGAGCCAGAAATATCTAAG	960
2470	QY	AAGACTGTTTAA	2481
961	Db	AAGACTGTTTAA	972

RESULT 11

RESOLUTION
AAAF98463

AAE98463
ID AAE98463 standard; cDNA: 2546 bp.XX
ID
AAFS

XX
AC

AC
AAFS
VV

XX
DE
EN
0000
7-7-77

DT 44

Human cpDNA clone CT585 1 sequence SEQ ID 150.

Human; secreted protein; nutrient; cytokine modulator; proliferation; differentiation; immune system modulator; tissue growth; chemotactic; haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss; haematopoiesis.

Homogamijens

WC300710'000-81

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

T-355, 3308-003388

(GEMY) GENETICS INST INC.

Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C, Verborg D, Treacy M, Bowman MP, Snoulding V, Acosta MC, MI.

U
C
✓
C
C
C
C
C
C
C
C
C
C

P-PSDB; AAB90727.

Isolated nucleic acids encoding polypeptides, useful for modulating e.g. cytokine and cell proliferation/differentiation activity, the immune system and hematopoiesis regulating activity.

Disclosures: Page 476-477: 557pp: English.

Human cDNA clones represented in AAF98374 - AAF98489 encode secreted proteins AAB90667 - AAB90750. The cDNA clones are isolated from various tissue types, and may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate protein expression. The polypeptides and nucleic acids may be used as nutrients or to modulate cytokine and cell proliferation/differentiation activity and may also be involved in modulation of the immune system. The cDNA sequences, proteins, their agonists and/or antagonists exhibit haematopoiesis-regulating activity; tissue growth activity; activin/inhibin activity; chemotactic/chemokinetic activity; haemostatic and thrombolytic activity; receptor/ligand activity; anti-inflammatory activity; haematopoiesis activity; cachetin/tumour suppressor activity; and/or tumour inhibition activity. Included in the invention are probes represented in AAF98490 - AAF98572 which are specific for the cDNA clones encoding the secreted proteins.

	90-91	89-90	88-89	87-88	86-87	85-86	84-85	83-84	82-83	81-82	80-81	79-80	78-79	77-78	76-77	75-76	74-75	73-74	72-73	71-72	70-71	69-70	68-69	67-68	66-67	65-66	64-65	63-64	62-63	61-62	60-61	59-60	58-59	57-58	56-57	55-56	54-55	53-54	52-53	51-52	50-51	49-50	48-49	47-48	46-47	45-46	44-45	43-44	42-43	41-42	40-41	39-40	38-39	37-38	36-37	35-36	34-35	33-34	32-33	31-32	30-31	29-30	28-29	27-28	26-27	25-26	24-25	23-24	22-23	21-22	20-21	19-20	18-19	17-18	16-17	15-16	14-15	13-14	12-13	11-12	10-11	9-10	8-9	7-8	6-7	5-6	4-5	3-4	2-3	1-2	0-1	Total	Average	Standard Deviation	Coefficient of Variation	Correlation Coefficient	Regression Line	Intercept	Slope	R-squared	F-statistic	p-value	t-statistic	df	Significance Level	Confidence Interval	Upper Bound	Lower Bound	Mean	Variance	Skewness	Kurtosis	Jarque-Bera Statistic	Ljung-Box Q	Pearson Chi-Square	Shapiro-Wilk W	Bartlett's Test	Mann-Whitney U	Kruskal-Wallis H	ANOVA F	Durbin-Watson d	Autocorrelation Function	Partial Autocorrelation Function	Box-Cox Lambda	Log Likelihood	AIC	BIC	Hausman Test	Nagelkerke R-squared	Cramer's V	Phi Coefficient	Odds Ratio	Relative Risk	Number of Events	Number at Risk	Life Table Estimate	Survival Probability	Hazard Rate	Incidence Rate	Prevalence Rate	Point Prevalence	Period Prevalence	Population Prevalence	Age-specific Prevalence	Gender-specific Prevalence	Ethnicity-specific Prevalence	Religion-specific Prevalence	Occupation-specific Prevalence	Education-specific Prevalence	Income-specific Prevalence	Marital Status-specific Prevalence	Health Insurance-specific Prevalence	Access to Care-specific Prevalence	Quality of Life-specific Prevalence	Functional Status-specific Prevalence	Mortality Rate-specific Prevalence	Morbidity Rate-specific Prevalence	Disability Rate-specific Prevalence	Longevity-specific Prevalence	Life Expectancy-specific Prevalence	Healthy Life Expectancy-specific Prevalence	Disability-adjusted Life Expectancy-specific Prevalence	Quality-adjusted Life Expectancy-specific Prevalence	Health-related Quality of Life-specific Prevalence	Physical Health-specific Prevalence	Mental Health-specific Prevalence	Social Health-specific Prevalence	Environmental Health-specific Prevalence	Overall Health-specific Prevalence	Summary Statistics	Notes	Comments	References	Source	Date	Page	Footnote	Disclaimer	Contact Information	Version	Revision History	Change Log	Update Schedule	Next Review Date	Review Status	Approval	Signature	Stamp	Seal	Mark	Initials	Abbreviations	Glossary	Index	Table of Contents	Appendix A	Appendix B	Appendix C	Appendix D	Appendix E	Appendix F	Appendix G	Appendix H	Appendix I	Appendix J	Appendix K	Appendix L	Appendix M	Appendix N	Appendix O	Appendix P	Appendix Q	Appendix R	Appendix S	Appendix T	Appendix U	Appendix V	Appendix W	Appendix X	Appendix Y	Appendix Z	Appendix AA	Appendix AB	Appendix AC	Appendix AD	Appendix AE	Appendix AF	Appendix AG	Appendix AH	Appendix AI	Appendix AJ	Appendix AK	Appendix AL	Appendix AM	Appendix AN	Appendix AO	Appendix AP	Appendix AQ	Appendix AR	Appendix AS	Appendix AT	Appendix AU	Appendix AV	Appendix AW	Appendix AX	Appendix AY	Appendix AZ	Appendix BA	Appendix BB	Appendix BC	Appendix BD	Appendix BE	Appendix BF	Appendix BG	Appendix BH	Appendix BI	Appendix BJ	Appendix BK	Appendix BL	Appendix BM	Appendix BN	Appendix BO	Appendix BP	Appendix BQ	Appendix BR	Appendix BS	Appendix BT	Appendix BU	Appendix BV	Appendix BW	Appendix BX	Appendix BY	Appendix BZ	Appendix CA	Appendix CB	Appendix CC	Appendix CD	Appendix CE	Appendix CF	Appendix CG	Appendix CH	Appendix CI	Appendix CJ	Appendix CK	Appendix CL	Appendix CM	Appendix CN	Appendix CO	Appendix CP	Appendix CQ	Appendix CR	Appendix CS	Appendix CT	Appendix CU	Appendix CV	Appendix CW	Appendix CX	Appendix CY	Appendix CZ	Appendix DA	Appendix DB	Appendix DC	Appendix DD	Appendix DE	Appendix DF	Appendix DG	Appendix DH	Appendix DI	Appendix DJ	Appendix DK	Appendix DL	Appendix DM	Appendix DN	Appendix DO	Appendix DP	Appendix DQ	Appendix DR	Appendix DS	Appendix DT	Appendix DU	Appendix DV	Appendix DW	Appendix DX	Appendix DY	Appendix DZ	Appendix EA	Appendix EB	Appendix EC	Appendix ED	Appendix EE	Appendix EF	Appendix EG	Appendix EH	Appendix EI	Appendix EJ	Appendix EK	Appendix EL	Appendix EM	Appendix EN	Appendix EO	Appendix EP	Appendix EQ	Appendix ER	Appendix ES	Appendix ET	Appendix EU	Appendix EV	Appendix EW	Appendix EX	Appendix EY	Appendix EZ	Appendix FA	Appendix FB	Appendix FC	Appendix FD	Appendix FE	Appendix FF	Appendix FG	Appendix FH	Appendix FI	Appendix FJ	Appendix FK	Appendix FL	Appendix FM	Appendix FN	Appendix FO	Appendix FP	Appendix FQ	Appendix FR	Appendix FS	Appendix FT	Appendix FU	Appendix FV	Appendix FW	Appendix FX	Appendix FY	Appendix FZ	Appendix GA	Appendix GB	Appendix GC	Appendix GD	Appendix GE	Appendix GF	Appendix GG	Appendix GH	Appendix GI	Appendix GJ	Appendix GK	Appendix GL	Appendix GM	Appendix GN	Appendix GO	Appendix GP	Appendix GQ	Appendix GR	Appendix GS	Appendix GT	Appendix GU	Appendix GV	Appendix GW	Appendix GX	Appendix GY	Appendix GZ	Appendix HA	Appendix HB	Appendix HC	Appendix HD	Appendix HE	Appendix HF	Appendix HG	Appendix HH	Appendix HI	Appendix HJ	Appendix HK	Appendix HL	Appendix HM	Appendix HN	Appendix HO	Appendix HP	Appendix HQ	Appendix HR	Appendix HS	Appendix HT	Appendix HU	Appendix HV	Appendix HW	Appendix HX	Appendix HY	Appendix HZ	Appendix IA	Appendix IB	Appendix IC	Appendix ID	Appendix IE	Appendix IF	Appendix IG	Appendix IH	Appendix II	Appendix IJ	Appendix IK	Appendix IL	Appendix IM	Appendix IN	Appendix IO	Appendix IP	Appendix IQ	Appendix IR	Appendix IS	Appendix IT	Appendix IU	Appendix IV	Appendix IW	Appendix IX	Appendix IY	Appendix IZ	Appendix JA	Appendix JB	Appendix JC	Appendix JD	Appendix JE	Appendix JF	Appendix JG	Appendix JH	Appendix JI	Appendix JJ	Appendix JK	Appendix JL	Appendix JM	Appendix JN	Appendix JO	Appendix JP	Appendix JQ	Appendix JR	Appendix JS	Appendix JT	Appendix JU	Appendix JV	Appendix JW	Appendix JX
--	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-------	---------	--------------------	--------------------------	-------------------------	-----------------	-----------	-------	-----------	-------------	---------	-------------	----	--------------------	---------------------	-------------	-------------	------	----------	----------	----------	-----------------------	-------------	--------------------	----------------	-----------------	----------------	------------------	---------	-----------------	--------------------------	----------------------------------	----------------	----------------	-----	-----	--------------	----------------------	------------	-----------------	------------	---------------	------------------	----------------	---------------------	----------------------	-------------	----------------	-----------------	------------------	-------------------	-----------------------	-------------------------	----------------------------	-------------------------------	------------------------------	--------------------------------	-------------------------------	----------------------------	------------------------------------	--------------------------------------	------------------------------------	-------------------------------------	---------------------------------------	------------------------------------	------------------------------------	-------------------------------------	-------------------------------	-------------------------------------	---	---	--	--	-------------------------------------	-----------------------------------	-----------------------------------	--	------------------------------------	--------------------	-------	----------	------------	--------	------	------	----------	------------	---------------------	---------	------------------	------------	-----------------	------------------	---------------	----------	-----------	-------	------	------	----------	---------------	----------	-------	-------------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

C
L
C
J
I
M

L
P
C

C
D
C

F
H
C

C
C
C
C

J
J
S

Local Similarity 100.0%; Pred. No. 0;

[illegible]

151V: AHHHGHACCAAGGAAH111G1A1GA1AAAGGCAGG1AAAG1TAAGGAAHCA19CCAG119AH 1509

1 AAAAGAAACCAAGGAAATTGTATGATAAGGCAGGTAAAGTGAGGAAACATGCAACTGAA 60

1570 CAGGAAAAAAGCTGAAGAGGGATTAGGCCCTAATATATAAAAGCATTGTCACCATGTTGATG 1629

61. CAGGAA⁺AA⁺AACTGAAGAGGGATTAGGCCCTAATAATAAAAGCATTTGTCA⁺CCATGTTGATG 120

1630 CTGATGCTATTGATGATGTTTGCTGTCCACTGTACCTGGGTCAAGCAATGCCCTACTCT 1689

121 CTGATGCTATTGATGATGTTTGCTGTCTCCACTGTACTGGGTCAAGCAATGCCCTACTCT 180

1690 AGTCCAAGTGTAGTCCTGGCCTCATACAATCATGATGGCACCAGGAATATCTTAGATGAT 1749

181 AGTCCAAAGTGTCCTGGCCTCATACAA TCATGATGGCACCAGGAATATCTTAGATGAT 240

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

•

QY 387 AGCATGGTATCCATAGGAAGATAGTAGGTGGTACTGTTTACCCAGGGTGGATGAAC 446
Db |||||
QY 121 AGCATGGTATCCATAGGAAGATAGTAGGTGGTACTGTTTACCCAGGGTGGATGAAC 180
Db |||||
QY 447 CGCTGGCTTATTCATTCGGATTTTAAATACATTAAGCAATACCTGTTACATAAGACGCT 506
Db |||||
QY 181 CGCTGGCTTATTCATTCGGATTTTAAATACATTAAGCAATACCTGTTACATAAGACGCT 240
Db |||||
QY 507 ATGTGTGTCTCTGACCAACTTTTAGGGCTTACATCTATATCTACTTCTCTGCTTAC 566
Db |||||
QY 241 ATGTGTGTCTCTGACCAACTTTTAGGGCTTACATCTATATCTACTTCTCTGCTTAC 300
Db |||||
QY 567 AAGAGAACTTTGGAAACCAAGGAGGAGCACTTTTAGCTGCTGTTTATTGCTATTGTACC 626
Db |||||
QY 301 AAGAGAACTTTGGAAACCAAGGAGGAGCACTTTTAGCTGCTGTTTATTGCTATTGTACC 360
Db |||||
QY 627 AGGCTACATATCTGGTTCAGTAGCTGATCTTGTATGATGAAGGCACTTCTATTGTC 686
Db |||||
QY 361 AGGCTACATATCTGGTTCAGTAGCTGATCTTGTATGATGAAGGCACTTCTATTGTC 420
Db |||||
QY 687 ACTTCAGTTCATACACTATTATCGGTAAATCTGTAAAACTGGGTCTAGTTTGTGGAC 746
Db |||||
QY 421 ACTTCAGTTCATACACTATTATCGGTAAATCTGTAAAACTGGGTCTAGTTTGTGGAC 480
Db |||||
QY 747 AATGTGCTGCTTATCTATATCTATATGCTCTGCTGGGTGTTATGTTAT 806
Db |||||
QY 481 AATGTGCTGCTTATCTATATCTATATGCTCTGCTGGGTGTTATGTTAT 540
Db |||||
QY 807 CATCAATCTTATTCATCGCATGATGTTGTTGTTTACTGATCAGAGATACAGCAAAAG 866
Db |||||
QY 541 CATCAATCTTATTCATCGCATGATGTTGTTGTTTACTGATCAGAGATACAGCAAAAG 600
Db |||||
QY 867 AGTCTACATAGCATATAGCACTTCTACATGTTGGTTTAAATATTAATCAATCAGATACC 926
Db |||||
QY 601 AGTCTACATAGCATATAGCACTTCTACATGTTGGTTTAAATATTAATCAATCAGATACC 660
Db |||||
QY 927 TTTGTGGATTCAGCAATCAGAAACAGTGAACATGCGAGTGCAGGTGCTTTTTC 986
Db |||||
QY 661 TTTGTGGATTCAGCAATCAGAAACAGTGAACATGCGAGTGCAGGTGCTTTTTC 720
Db |||||
QY 987 ATTGTGCAAGCTTATGCTTCTTCTGAGTATCTGAGAGCCGATTAAACAAACAGAGTT 1046
Db |||||
QY 721 ATTGTGCAAGCTTATGCTTCTTCTGAGTATCTGAGAGCCGATTAAACAAACAGAGTT 780
Db |||||
QY 1047 CCAGACCTTTCTTTTGGGTGATCAGTACAGTGAACATGCGAGTGCAGGTGCTTTTTC 1106
Db |||||
QY 781 CCAGACCTTTCTTTTGGGTGATCAGTACAGTGAACATGCGAGTGCAGGTGCTTTTTC 840
Db |||||
QY 1107 CTATTGACTTATACAGTTACATTCACCATGAGTGGCAGGTTTATTCTATTGTTGGA 1166
Db |||||
QY 841 CTATTGACTTATACAGTTACATTCACCATGAGTGGCAGGTTTATTCTATTGTTGGA 900
Db |||||
QY 1167 TACTGGGTATGCAAAATACATTCCAATTTGATGATGATGATGATGATGATGATGATGAT 1226
Db |||||
QY 901 TACTGGGTATGCAAAATACATTCCAATTTGATGATGATGATGATGATGATGATGATGAT 960
Db |||||
QY 1227 GACTGGGTGCTTTCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1286
Db |||||
QY 961 GACTGGGTGCTTTCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
Db |||||
QY 1287 TTGTTCTGCTCAAAATACATTCAGATGATGATGATGATGATGATGATGATGATGATGAT 1320
Db |||||
QY 1021 TTGTTCTGCTCAAAATACATTCAGATGATGATGATGATGATGATGATGATGATGATGAT 1054
Db |||||

RESULT 9

ADN60663

ID ADN60663 standard; cDNA; 1543 BP.

XX AC ADN60663;

XX AC ADN60663;

DT 01-JUL-2004 (first entry)

XX

Human secreted polynucleotide #123.

Human; secreted polynucleotide; gene; ss; autoimmune disease; rheumatoid arthritis; hyperproliferative disorder; neoplasm; breast; liver; cardiovascular disorder; cardiac arrest; cerebrovascular disorder; cerebral ischaemia; nervous system disorder; Alzheimer's disease; bacterial infection; viral infection; fungal infection; ocular disorder; corneal infection; wound healing; epithelial cell proliferation; skin aging; sunburn; organ transplantation; tissue regeneration; chemotaxis; food additive; food preservative; fat content; vitamin; mineral.

Homo sapiens.

US2004038277-A1.

26-FEB-2004.

18-JUL-2003; 2003US-00621401.

07-NOV-1997; 97US-0064900P.
07-NOV-1997; 97US-0064908P.
07-NOV-1997; 97US-0064911P.
07-NOV-1997; 97US-0064912P.
07-NOV-1997; 97US-0064983P.
07-NOV-1997; 97US-0064984P.
07-NOV-1997; 97US-0064985P.
07-NOV-1997; 97US-0064987P.
07-NOV-1997; 97US-0064988P.
07-NOV-1997; 97US-0066089P.
07-NOV-1997; 97US-0066090P.
07-NOV-1997; 97US-0066094P.
07-NOV-1997; 97US-0066095P.
07-NOV-1997; 97US-0066100P.
04-NOV-1996; 98WO-00023435.
05-MAY-1999; 99US-000305735.
13-OCT-2000; 2000US-0239893P.
28-MAR-2001; 2001US-00818683.
12-OCT-2001; 2001US-00974879.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Feng P, Ruben SM, Ebner R, Olsen H, Ni J, Wei Y;
Soppet DR, Moore PA, Kyaw H, Laflaur DW, Shi Y, Janat P;
Endress GA, Carter KC, Birse CE;

WPI; 2004-203215/19.

P-PSDB; ADN60791.

New nucleic acid molecule encoding one of 125 human secreted proteins for diagnosing, preventing, treating or ameliorating medical conditions and used as food additives or preservatives.

Claim 1: SEQ ID NO 133; 531pp; English.

The invention relates to isolated nucleic acid molecules encoding human secreted proteins. The polynucleotides and polypeptides are used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the polypeptides can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases (e.g. rheumatoid arthritis), hyperproliferative disorders (e.g. neoplasms of the breast or liver), cardiovascular disorders (e.g. cardiac arrest), cerebrovascular disorders (e.g. cerebral ischaemia), nervous system disorders (e.g. Alzheimer's disease), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infection). The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as food additives or preservatives to increase or decrease storage capabilities and fat content, as vitamins, as minerals and as other

QY 567 AAGAGAACTTTGGAAACCAAGGACAGAGACTTTTGTGCTGCTTTTATGCTATGTTAC 626
 DB 301 AAGAGAACTTTGGAAACCAAGGACAGAGACTTTTGTGCTGCTTTTATGCTATGTTAC 360
 QY 627 AGGCTACATATCTCGGTGAGTGTGATCTCTTGTATGATGAAGGATGCTATTTTGGC 686
 DB 361 AGGCTACATATCTCGGTGAGTGTGATCTCTTGTATGATGAAGGATGCTATTTTGGC 420
 QY 687 ACTTCAGTTCACATATGTTGGTAAATCTGTAATGCTGAGTGTGCTATTTTGGAC 746
 DB 421 ACTTCAGTTCACATATGTTGGTAAATCTGTAATGCTGAGTGTGCTATTTTGGAC 480
 QY 747 AATGTGCTGCTTATCTTATTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 806
 DB 481 AATGTGCTGCTTATCTTATTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 807 CATCAATCTTATTCCTGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 866
 DB 541 CATCAATCTTATTCCTGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 600
 QY 867 AGTCTACATAGCATATGACCTTCTACATGTTGCTGTTGCTGTTGCTGTTGCTGTTG 926
 DB 601 AGTCTACATAGCATATGACCTTCTACATGTTGCTGTTGCTGTTGCTGTTGCTGTTG 660
 QY 927 TTTTGTGGATTCAGGCAATCAGAACAGTGAACATGCGAGCTGCGAGTGTCTTGGC 986
 DB 661 TTTTGTGGATTCAGGCAATCAGAACAGTGAACATGCGAGCTGCGAGTGTCTTGGC 720
 QY 987 ATTGTGCAAGCTTATGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1046
 DB 721 ATTGTGCAAGCTTATGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 QY 1047 CCAGACCTTTCTTTTGGGTGATCACTAGCTGAGTGTGCTGCTGCTGCTGCTGCTGCT 1106
 DB 781 CCAGACCTTTCTTTTGGGTGATCACTAGCTGAGTGTGCTGCTGCTGCTGCTGCTGCT 840
 QY 1107 CTATTTGACTTATACAGTTCATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1166
 DB 841 CTATTTGACTTATACAGTTCATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 QY 1167 TACTGGGTATCAAAATACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1226
 DB 901 TACTGGGTATCAAAATACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 960
 QY 1227 GACTTGGGTGCTTTCTTTCTTTGATCTACATATTCCTGATGATGATGATGATGATGAT 1286
 DB 961 GACTTGGGTGCTTTCTTTCTTTGATCTACATATTCCTGATGATGATGATGATGATGAT 1020
 QY 1287 TTGGTTCGATCAAAATATCAAGATGAAGA 1320
 DB 1021 TTGGTTCGATCAAAATATCAAGATGAAGA 1054

RESULT 8

ADG78372

ID ADG78372 standard; cDNA; 1543 BP.

AC ADG78372;

XX ADG78372;

DT 11-MAR-2004 (first entry)

XX Human secreted protein cDNA #123.

DE Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

XX Human; secreted protein; gene; ss; immune disorder;

OS Homo sapiens.
 XX US2003211472-A1.
 PN 13-NOV-2003.
 PD 28-MAR-2001; 2001US-00818683.
 PF 05-MAY-1999; 99US-00305736.
 XX (FENG/) FENG P.
 XX (RUBEN/) RUBEN S M.
 XX (ROSE/) ROSEN C A.
 XX (EBNER/) EBNER R.
 XX (OLSEN/) OLSEN H S.
 XX (NIJ/) NI J.
 XX (WEI/) WEI Y.
 XX (SOPP/) SOPPET D R.
 XX (MOORE/) MOORE P A.
 XX (KYAW/) KYAW H.
 XX (LAFLEUR/) LAFLEUR D W.
 XX (SHI/) SHI Y.
 XX (JANAT/) JANAT F.
 XX (ENDRESS/) ENDRESS G A.
 XX (CARTER/) CARTER K C.
 XX Feng P, Ruben SM, Rosen CA, Ebner R, Olsen HS, Ni J, Wei Y;
 PI Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;
 PI Endress GA, Carter KC;
 XX WPI: 2004-051575/05.
 DR P-PSDB; ADG78500.

XX New secreted nucleic acid for diagnosing, preventing or treating diseases
 PT associated with aberrant expression or activity of the polypeptide it
 PT encodes, e.g. cancer, human immunodeficiency virus, Parkinson's disease,
 PT or diabetes.

PS Claim 1; SEQ ID NO 133; 377pp; English.

CC The invention also relates to human secreted proteins and the nucleic
 CC acids encoding them. The proteins and nucleic acids are useful in
 CC diagnosing, preventing, prognosing or treating diseases or disorders
 CC associated with aberrant expression and/or activity of the secreted
 CC proteins, such as immune disorders, haematopoietic disorders,
 CC hyperproliferative disorders, infectious diseases or inflammatory
 CC disorders. In particular, the diseases or disorders are HIV, anaemia,
 CC thrombocytopenia, bleeding, stroke, myocardial infarction, Addison's
 CC disease, rheumatoid arthritis, dermatitis, Grave's disease, multiple
 CC sclerosis, glomerulonephritis, diabetes, graft-versus-host disease,
 CC inflammatory bowel disease, cancer, bacterial infections, viral
 CC infections, parasitic infections, osteoporosis, wounds, atherosclerosis,
 CC Alzheimer's disease, Parkinson's disease or ulcers. The sequences may
 CC also be used as food additives or preservatives, or for modulating
 CC mammalian mental or physical characteristics. The nucleic acids are also
 CC used in chromosome mapping, in forensic biology or as molecular weight
 CC markers. This sequence represents cDNA encoding a human secreted protein
 CC of the invention.

XX Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;

Query Match 42.5%; Score 1054; DB 12; Length 1543;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 CTTCCCGCTCATCCGCTTCGAAAGCATCATCCAGAGTTCGACCCGCTGTTAACTATAG 326
 DB 1 CTTCCCGCTCATCCGCTTCGAAAGCATCATCCAGAGTTCGACCCGCTGTTAACTATAG 60
 QY 327 ATCAACACATCATCTTGATCTCATGGTCTCATGAATTTTAAATGTTGATGAAG 386
 DB 61 ATCAACACATCATCTTGATCTCATGGTCTCATGAATTTTAAATGTTGATGAAG 120

```
Db 541 CATCACTTATTCACATGCAATGATTTGTTGTTTACTGATGACAGATACAGCAAG 600
QY 867 AGTCTACATAGCATATAGCACTTTCTACATGTTGGGTTTAAATATCAATCAGATACC 926
Db 601 AGTCTACATAGCATATAGCACTTTCTACATGTTGGGTTTAAATATCAATCAGATACC 660
QY 927 TTTTGTGGGATCCAGGCAATCAGAACAAAGTGAACACATGGCAGCTGCGAGTGTCTTTCG 986
Db 661 TTTTGTGGGATCCAGGCAATCAGAACAAAGTGAACACATGGCAGCTGCGAGTGTCTTTCG 720
QY 987 ATTGCTGCAAGCTTATGCTTTCTTTCGAGTATCTGAGAGCCGATTAACAAAACAGAGTT 1046
Db 721 ATTGCTGCAAGCTTATGCTTTCTTTCGAGTATCTGAGAGCCGATTAACAAAACAGAGTT 780
QY 1047 CCAGACCCCTTTCTTTTGGGTGATCACTAGCTGCGAGTGTCTTCTTCTAGTGTCTAT 1106
Db 781 CCAGACCCCTTTCTTTTGGGTGATCACTAGCTGCGAGTGTCTTCTTCTAGTGTCTAT 840
QY 1107 CTATTTGACTTATACAGGTTACATTCACCATGAGTGGCAGGTTTATTCATTTGGGA 1166
Db 841 CTATTTGACTTATACAGGTTACATTCACCATGAGTGGCAGGTTTATTCATTTGGGA 900
QY 1167 TACTGGGTATGCAAAAATACACATTCCTCAATTTATTCATCATGCTGTGAGCATCAACCTAC 1226
Db 901 TACTGGGTATGCAAAAATACACATTCCTCAATTTATTCATCATGCTGTGAGCATCAACCTAC 960
QY 1227 GACTGGGTGCTTCTTCTTCTTGTATACATATTTCTGTATGCTACCTCCAGCAGGCT 1286
Db 961 GACTGGGTGCTTCTTCTTCTTGTATACATATTTCTGTATGCTACCTCCAGCAGGCT 1020
QY 1287 TTGGTTCTGCTCAAAAATATCAACGATGAAAGA 1320
Db 1021 TTGGTTCTGCTCAAAAATATCAACGATGAAAGA 1054
```

RESULT 7

```
ACD18981
ID ACD18981 standard; cDNA; 1543 BP.
XX AC ACD18981;
XX DT 21-AUG-2003 (first entry)
XX DE Novel human secreted protein cDNA #123.
XX KW Human; secreted protein; cytosolic; neuroprotective; hepatotropic;
KW Gene therapy; cancer; liver disorder; hepatitis; neural disorder;
KW Alzheimer's disease; gene; ss.
XX OS Homo sapiens.
XX PN US2003028003-A1.
XX PD 06-FEB-2003.
XX PF 12-OCT-2001; 2001US-00974879.
XX 07-NOV-1997; 97US-0064900P.
PR 07-NOV-1997; 97US-0064908P.
PR 07-NOV-1997; 97US-0064911P.
PR 07-NOV-1997; 97US-0064912P.
PR 07-NOV-1997; 97US-0064983P.
PR 07-NOV-1997; 97US-0064984P.
PR 07-NOV-1997; 97US-0064985P.
PR 07-NOV-1997; 97US-0064987P.
PR 17-NOV-1997; 97US-0066088P.
PR 17-NOV-1997; 97US-0066089P.
PR 17-NOV-1997; 97US-0066090P.
PR 17-NOV-1997; 97US-0066094P.
PR 17-NOV-1997; 97US-0066095P.
PR 17-NOV-1997; 97US-0066100P.
PR 04-NOV-1998; 98WO-US023435.
```

```
PR 05-MAY-1999; 99US-00305736.
PR 13-OCT-2000; 2000US-0239893P.
PR 28-MAR-2001; 2001US-00818683.
XX (ROSE/) ROSEN C A.
PA (FENG/) FENG P.
PA (RUBEN/) RUBEN S M.
PA (EBNER/) EBNER R.
PA (OLSEN/) OLSEN H S.
PA (NIJJ/) NI J.
PA (WEI/) WEI Y.
PA (SOPP/) SOPPET D R.
PA (MOORE/) MOORE P A.
PA (KYAW/) KYAW H.
PA (LAFLEUR/) LAFLEUR D W.
PA (SHI/) SHI Y.
PA (JANAT/) JANAT F.
PA (ENDRESS/) ENDRESS G A.
PA (CASTER/) CASTER K C.
PA (BIRSE/) BIRSE C E.
XX PI Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;
PI Endress GA, Carter KC, Birse CE;
XX WPI: 2003-479549/45.
XX P-PSDB; ABO14034.
XX New nucleic acid molecule, useful for preparing a medicament for
XX preventing, treating or ameliorating a medical condition e.g., cancer,
XX liver disorders such as hepatitis or neural disorders such as Alzheimer's
XX disease.
XX Claim 1; Page 349; 496pp; English.
XX The invention describes a new isolated nucleic acid molecule comprising a
XX sequence having at least 95% identity with a sequence comprising: (a) a
XX polynucleotide (PN) fragment of a sequence comprising 420-3435 bp, or its
XX allelic variant; (b) a PN fragment of the cDNA sequence; (c) a PN
XX sequence encoding a polypeptide, or its fragment, domain, epitope or
XX species homologue; or (d) a PN that hybridises under stringent conditions
XX to any one of the sequences of (A)-(C). The nucleic acid is useful for
XX preparing a medicament for preventing, treating or ameliorating a medical
XX condition e.g., cancer, liver disorders such as hepatitis or neural
XX disorders such as Alzheimer's disease. This sequence encodes a novel
XX human secreted protein
XX SQ Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;
```

```
Query Match 42.5%; Score 1054; DB 8; Length 1543;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 267. CTTGGCGGTTCATCCGCTTGGAAAGCATCATCCAGAGTTCGACCCGTTTAACTATAG 326
Db 1. CTTGGCGGTTCATCCGCTTGGAAAGCATCATCCAGAGTTCGACCCGTTTAACTATAG 60
QY 327. ATCAACACATCATCTTGCATCTCATGGTTCATGAATTTTAAATGGTTGATGAAG 386
Db 61. ATCAACACATCATCTTGCATCTCATGGTTCATGAATTTTAAATGGTTGATGAAG 120
QY 387. AGCATGGTATCCAGTAGGAAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATAAC 446
Db 121. AGCATGGTATCCAGTAGGAAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATAAC 180
QY 447. CGCTGGCCTTATTCATTTGATTTTAAATACATTAACATCTCTTCAATAAGACAGCT 506
Db 181. CGCTGGCCTTATTCATTTGATTTTAAATACATTAACATCTCTTCAATAAGACAGCT 240
QY 507. ATGTGTGTTCCCTTGGACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 566
Db 241. ATGTGTGTTCCCTTGGACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 300
```

Db 601 AGTCTACATAGCATATAGCACTTCTTCTACATTGTGGGTTTAAATATTATATCAATGCAGATACC 660
 QY 927 TTTTGTGGGATCCAGCCATCAGAACAGTGAACATAGGCGAGGTGCTTTTGC 986
 Db 661 TTTTGTGGGATCCAGCCATCAGAACAGTGAACATAGGCGAGGTGCTTTTGC 720
 QY 987 ATTGTGCAAGCTTATGCTTTCTTGTGAGTATCTGAGAGACCGGATTAAACAAAACAGAGTT 1046
 Db 721 ATTGTGCAAGCTTATGCTTTCTTGTGAGTATCTGAGAGACCGGATTAAACAAAACAGAGTT 780
 QY 1047 CAGAGCCCTTTCTTTTGTGGTATCACTAGCTGAGGTGCTGCTTGTAGTGTAT 1106
 Db 781 CAGAGCCCTTTCTTTTGTGGTATCACTAGCTGAGGTGCTGCTTGTAGTGTAT 840
 QY 1107 CTATTGTACTTATACAGGTTACATTCACCATGAGTGGCAGGTTTATTTCATTGTGGGA 1166
 Db 841 CTATTGTACTTATACAGGTTACATTCACCATGAGTGGCAGGTTTATTTCATTGTGGGA 900
 QY 1167 TACTGGGTATCAAAATACATTCACCAATTATTGATCAGTGTCTGAGCATCAACCTAC 1226
 Db 901 TACTGGGTATCAAAATACATTCACCAATTATTGATCAGTGTCTGAGCATCAACCTAC 960
 QY 1227 GACTGGGTGCTTTCTTCTTTGTATGATACATATCTTGTATGATCTCCAGAGGCT 1286
 Db 961 GACTGGGTGCTTTCTTCTTTGTATGATACATATCTTGTATGATCTCCAGAGGCT 1020
 QY 1287 TTGGTTCTGCATCAAAATATCAACGATGAAGA 1320
 Db 1021 TTGGTTCTGCATCAAAATATCAACGATGAAGA 1054

RESULT 6

AA85055

ID AA85055 standard; DNA; 1543 BP.

AC AA85055;

XX 30-JUL-1999 (first entry)

DE Human secreted protein gene No. 123.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

OS

XX WO924836-A1.

PN

XX 20-MAY-1999.

PD

XX 04-NOV-1996; 98WO-US023435.

PF

XX 07-NOV-1997; 97US-0064900P.

PR

XX 07-NOV-1997; 97US-0064908P.

PR

XX 07-NOV-1997; 97US-0064911P.

PR

XX 07-NOV-1997; 97US-0064912P.

PR

XX 07-NOV-1997; 97US-0064983P.

PR

XX 07-NOV-1997; 97US-0064984P.

PR

XX 07-NOV-1997; 97US-0064985P.

PR

XX 07-NOV-1997; 97US-0064987P.

PR

XX 07-NOV-1997; 97US-0064988P.

PR

XX 07-NOV-1997; 97US-0066008P.

PA (HUMA-) HUMAN GENOME SCI INC.
 PI Peng P, Carter KC, Endress CA, Rosen CA, Ruben SM, Janat F;
 PI Ni J, Wei Y, Moore PA, Soppet DR, Kyaw H, Lafleur DW, Olsen HS;
 PI Shi Y, Ebner R;
 XX
 DR WPI; 1999-337740/28.
 DR P-PSDB; AAY27689, AAY27918, AAY27919, AAY27920, AAY27921, AAY27922,
 DR AAY27923.
 XX
 PT New human secreted proteins and coding sequences useful for treating
 PT disorders of the immune system and hyperproliferative disorders.
 PS
 PS Claim 1: Page 347; 507pp; English.
 XX
 CC This sequence represents a nucleic acid molecule which encodes a secreted
 CC human protein. The gene number is given in the descriptor line. The gene
 CC can be used to generate fusion proteins by linking to the gene to a human
 CC immunoglobulin Fc portion (e.g. AAX84924) for increasing the stability of
 CC the fused protein as compared to the human protein only. The invention
 CC relates to 125 novel genes and their fragments (nucleic acid sequences;
 CC AAX84933-X85057; amino acid sequences AAY27567-127933) which are useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. Also, pathological conditions can be diagnosed
 CC by determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the 125 polynucleotides, based on
 CC which tissues they are most highly expressed in (see AAX84933 for
 CC described uses)

XX SQ Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;

Query Match 42.5%; Score 1054; DB 2; Length 1543;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 CTTCCGCGTCATCGCTCGAAGCATCATCCAGAGTTCGACCCGCTGTTAACTATAG 326

Db 1 CTTCCGCGTCATCGCTCGAAGCATCATCCAGAGTTCGACCCGCTGTTAACTATAG 60

QY 327 ATCAACACATCATCTGCACTCTCATGGTCTCTATGAATTTTAAATGGTTGATGAAG 386

Db 61 ATCAACACATCATCTGCACTCTCATGGTCTCTATGAATTTTAAATGGTTGATGAAG 120

QY 387 AGCTGGTATCCATAGGAAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATAAC 446

Db 121 AGCTGGTATCCATAGGAAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATAAC 180

QY 447 CGCTGGCTTATTCATTGGATTTTAAATACATTAACATCACTGTTTCAATAAGACGCT 506

Db 181 CGCTGGCTTATTCATTGGATTTTAAATACATTAACATCACTGTTTCAATAAGACGCT 240

QY 507 ATGTGTGTTCTTGCACCAACTTTTAGCGGCTTACATCTATCTACTTCTCTGCTTAC 566

Db 241 ATGTGTGTTCTTGCACCAACTTTTAGCGGCTTACATCTATCTACTTCTCTGCTTAC 300

QY 567 AAGAGAACTTTGGAAACCAAGGAGGAGACTTTTAGCTGCTGTTTATTCGTTATTTTAC 626

Db 301 AAGAGAACTTTGGAAACCAAGGAGGAGACTTTTAGCTGCTGTTTATTCGTTATTTTAC 360

QY 627 AGGCTACATATCTCGTTCAGTGTGATCCTTTGATAATGAAGGATGCTATTTTTCG 686

Db 361 AGGCTACATATCTCGTTCAGTGTGATCCTTTGATAATGAAGGATGCTATTTTTCG 420

QY 687 ACTTCAGTTCACATCTATTTTATGGTAAATCTGTAATACTGGGTGAGTTTGTGGAC 746

Db 421 ACTTCAGTTCACATCTATTTTATGGTAAATCTGTAATACTGGGTGAGTTTGTGGAC 480

QY 747 AATGTGCTGCTTATCTATCTATATGCTCTGCTTGGGGTGGTTATCTATTTAT 806

Db 481 AATGTGCTGCTTATCTATCTATATGCTCTGCTTGGGGTGGTTATCTATTTAT 540

QY 807 CATCAATCTTATCCACTGTCATGTAATTTGTGTTGTTACTGATGCAGATACAGCAAG 866

Db 628 GCTGCTGTTTATTTGCTATTTGTTACAGGCTACATATCTCGGTAGTGGATCCCTTT 687
Qy 661 GATATGAGGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 720
Db 688 GATAATGAGGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 747
Qy 721 GTAAAACTGGGTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 780
Db 748 GTAAAACTGGGTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 807
Qy 781 TCTGCTGCTGGGTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 840
Db 808 TCTGCTGCTGGGTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 867
Qy 841 TTACTGATGAGGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 900
Db 868 TTACTGATGAGGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 927
Qy 901 GGTTTAAATATTAATCAATGAGGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 960
Db 928 GGTTTAAATATTAATCAATGAGGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 987
Qy 961 CATATGAGGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 1020
Db 988 CATATGAGGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 1047
Qy 1021 AGAGACCGATTAATCAATGAGGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 1080
Db 1048 AGAGACCGATTAATCAATGAGGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 1107
Qy 1081 GCAGTGCTGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 1140
Db 1108 GCAGTGCTGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 1167
Qy 1141 AGTGACGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 1200
Db 1168 AGTGACGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 1227
Qy 1201 GCATGAGTGTGAGCATCAACCTTACAGTTCAGTTCACATACATATTTATGTTGGTAAATCT 1260
Db 1228 GCATGAGTGTGAGCATCAACCTTACAGTTCAGTTCACATACATATTTATGTTGGTAAATCT 1287
Qy 1261 CTGTATGCTGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 1320
Db 1288 CTGTATGCTGCTGCTATTTTGGACCTTCAGTTCACATACATATTTATGTTGGTAAATCT 1347
Qy 1321 GTATTTG 1327
Db 1348 GTATTTG 1354

RESULT 4
AAH33264
ID AAH33264 standard; cDNA, 1209 BP.
XX
AC AAH33264;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:320.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma, ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX

PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
DR P-PSDB; AAG73833.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
PS Claim 1; Page 2436; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patient's own production of P. Additionally, N may be used
XX to produce the colon cancer-associated Ps, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins. N and P
XX can be used in the prevention, diagnosis and treatment of colorectal
XX carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
XX sequences used in the exemplification of the present invention. N.B.
XX Pages 866 to 882 and page 7053 of the sequence listing were missing at
XX time of publication, meaning no sequences are present for SEQ ID NO:1027
XX to 1052, 7921 and 7922
SQ Sequence 1209 BP; 312 A; 222 C; 235 G; 438 T; 0 U; 2 Other;
Query Match 42.5%; Score 1054; DB 4; Length 1209;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 267 CTTGCGCTGATCCCTTCGAAAGCATCATCCAGAGTTCGACCGGTGTTAACTATAG 326
Db 1 CTTGCGCTGATCCCTTCGAAAGCATCATCCAGAGTTCGACCGGTGTTAACTATAG 60
Qy 327 ATCAACACATCATCTTCATCTCTCATGGTCTCATGATTTTAAATTTGTTTCAATGAAAG 386
Db 61 ATCAACACATCATCTTCATCTCTCATGGTCTCATGATTTTAAATTTGTTTCAATGAAAG 120
Qy 387 AGCATGGTATCCACTAGGAAGANTAGTGGTGGTACCTTACCAGGTTGATGATAAC 446
Db 121 AGCATGGTATCCACTAGGAAGANTAGTGGTGGTACCTTACCAGGTTGATGATAAC 180
Qy 447 CGCTGGCTTATTCATTTGATTTTAAATATACATTAACATTAACATTAACATTAACATTAAC 506
Db 181 CGCTGGCTTATTCATTTGATTTTAAATATACATTAACATTAACATTAACATTAACATTAAC 240
Qy 507 ATGTGTGTTCTTCGACCAACTTTTACGGCTTACATCTATATCTACTTCTCTGCTTAC 566
Db 241 ATGTGTGTTCTTCGACCAACTTTTACGGCTTACATCTATATCTACTTCTCTGCTTAC 300
Qy 567 AAGAGAACTTTTGGAAACCAAGGAGGAGGAGCTTTTACGCTGCTTGTATTTATGCTATTGACC 626
Db 301 AAGAGAACTTTTGGAAACCAAGGAGGAGGAGCTTTTACGCTGCTTGTATTTATGCTATTGACC 360
Qy 627 AGGTACATATCTCGGTGAGTGGTGGTCTTTGATTAAGTAAAGGAGGAGGAGGAGGAGGAGG 686
Db 361 AGGTACATATCTCGGTGAGTGGTGGTCTTTGATTAAGTAAAGGAGGAGGAGGAGGAGGAGG 420
Qy 687 ACTTCAGTTTCACATATCTTTTGGTAAATCTGTAAATCTGTAAATCTGTAAATCTGTAAATCT 746
Db 421 ACTTCAGTTTCACATATCTTTTGGTAAATCTGTAAATCTGTAAATCTGTAAATCTGTAAATCT 480
Qy 747 AATGTCGCTGCTTATCTTATCTATATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 806

721	Db	GTAAAAA	CTGGG	CAGT	TTTTTGG	CAATG	TGCTG	CTGCT	TATCCT	ATTTCT	ATATG	CGTC	780
781	Qy	TCTGCT	TTGGGG	TGTTAT	GTATTAT	CATCA	ATCTTAT	TCCAC	TGCGAT	TTATTTG	TGCTTG	840	
781	Db	TCTGCT	TTGGGG	TGTTAT	GTATTAT	CATCA	ATCTTAT	TCCAC	TGCGAT	TTATTTG	TGCTTG	840	
841	Qy	TTACTG	ATGCAG	AGATAC	AGAAAG	AGTCT	ACATAG	CATATAC	GAC	CTTCTT	CATTTG	900	
841	Db	TTACTG	ATGCAG	AGATAC	AGAAAG	AGTCT	ACATAG	CATATAC	GAC	CTTCTT	CATTTG	900	
901	Qy	GGTTTA	TATTTAT	CAATG	CAGAT	TACCTTTG	TGGGAT	TCCAG	CCAAATC	CAGAAC	CAAGTAA	960	
901	Db	GGTTTA	TATTTAT	CAATG	CAGAT	TACCTTTG	TGGGAT	TCCAG	CCAAATC	CAGAAC	CAAGTAA	960	
961	Qy	CACATG	CAGCTG	GAGGTG	CTTTTGC	ATTG	CTGCA	AGCTTAT	TGCTTCT	TTCG	CAGTATCTG	1020	

1021	QY	AGAGACCGANTTAA	CAAAAACAAGAGTTC	CGAGACCCCTTTTCTTTTGGGTGATCACTAGCT	1080
1021	DB	AGAGACCGAGTAA	CAAAAACAAGAGTTC	CGAGACCCCTTTTCTTTTGGGTGATCACTAGCT	1080
1081	QY	CGAGTCTGCTGTTCTT	AGTTCATCTATTTGACTTAT	CAGGTTACANTGCACCACTGG	1140
1081	DB	CGAGTCTGCTGTTCTT	AGTTCATCTATTTGACTTAT	CAGGTTACANTGCACCACTGG	1140
1141	QY	AGTGGCAGGTTTTAT	TCATTTGTGGGATCTGGGTA	TGCAAAAATACACATTCGAATTAAT	1200
1141	DB	AGTGGCAGGTTTTAT	TCATTTGTGGGATCTGGGTA	TGCAAAAATACACATTCGAATTAAT	1200
1201	QY	GCATCAGTGTCTGAGCAT	CAAACTCAGACTTGGGTTG	TCTTCTTTGATCTACATATT	1260
1201	DB	GCATCAGTGTCTGAGCAT	CAAACTCAGACTTGGGTTG	TCTTCTTTGATCTACATATT	1260
1261	QY	CTTGATGTACTCTCCAG	CAGGCGCTTTGGTTCTGCAT	CAAAAATACACAGTGAAGA	1320
1261	DB	CTTGATGTACTCTCCAG	CAGGCGCTTTGGTTCTGCAT	CAAAAATACACAGTGAAGA	1320
1321	QY	GTAATTTGTTGCTCTAT	ATATGCAATCAGTGTCTACT	TGCTGGAGTGATGGTSCGACTG	1380
1321	DB	GTAATTTGTTGCTCTAT	ATATGCAATCAGTGTCTACT	TGCTGGAGTGATGGTSCGACTG	1380
1381	QY	ATGTTGACTTTTGACT	CCAGTCTGTGTATGCTCTG	CAATTCCTTTTCAAAATGTTTTT	1440
1381	DB	ATGTTGACTTTTGACT	CCAGTCTGTGTATGCTCTG	CAATTCCTTTTCAAAATGTTTTT	1440
1441	QY	GAGCACAATTTGGGGGAT	GACATGAAAAGGGAATA	TCCACCTGTGGAGGACAGCAGTGAT	1500
1441	DB	GAGCACAATTTGGGGGAT	GACATGAAAAGGGAATA	TCCACCTGTGGAGGACAGCAGTGAT	1500
1501	QY	GAGGATGACAAAGAAAC	CAAGGAATTTGTATGATAG	GCGATTAAGTGAAGGAACAT	1560
1501	DB	GAGGATGACAAAGAAAC	CAAGGAATTTGTATGATAG	GCGATTAAGTGAAGGAACAT	1560
1561	QY	GCAACTGAAACAGGAAAAA	CTGAAGAGGATTAGGCCCT	TATATAAAAGCATTTGTCAAC	1620
1561	DB	GCAACTGAAACAGGAAAAA	CTGAAGAGGATTAGGCCCT	TATATAAAAGCATTTGTCAAC	1620
1621	QY	ATGTTGATGCTGATCTAT	TGATGATGTTTGTCTGTCC	ACTGTACTGGTCAACAGCAAT	1680
1621	DB	ATGTTGATGCTGATCTAT	TGATGATGTTTGTCTGTCC	ACTGTACTGGTCAACAGCAAT	1680
1681	QY	GCCTACTCTAGTCCAAAG	TGAGTGTAGTCTTGGGCT	CATCAATCATGATGGCACCGAATATC	1740
1681	DB	GCCTACTCTAGTCCAAAG	TGAGTGTAGTCTTGGGCT	CATCAATCATGATGGCACCGAATATC	1740
1741	QY	TTAGATGATTTTAGAGAG	CGTTACTTTTGGCTTAAG	CGCAAAATACAGATGAACATGACGA	1800
1741	DB	TTAGATGATTTTAGAGAG	CGTTACTTTTGGCTTAAG	CGCAAAATACAGATGAACATGACGA	1800
1801	QY	GTAATGTCTTTGGTGGG	ATTATGCTTATCAGATAG	CTGGGAATGGCTTAATAGAACTACGTTG	1860

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 15, 2004, 00:53:33 ; Search time 1667 Seconds
(without alignments)
7812.721 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagccctcgcccc.....tatctaagaagactgttaa 2481

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0
4134886 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_23Sep04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002as.*
7: Geneseq2002bs.*
8: Geneseq2003as.*
9: Geneseq2003bs.*
10: Geneseq2003cs.*
11: Geneseq2003ds.*
12: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2481	100.0	2481	10 ADD94783	Adg94783 Human SIM
2	1276	51.4	1664	AAK94164	AAK94164 Human ful
3	1276	51.4	1664	ADL30661	ADL30661 Full leng
4	1054	42.5	1209	AAH33284	AAH33284 Human col
5	1054	42.5	1209	ABL89850	ABL89850 Human pol
6	1054	42.5	1543	2 AAX85055	AAX85055 Human sec
7	1054	42.5	1543	8 ACD18981	ACD18981 Novel hum
8	1054	42.5	1543	12 ADG78372	Adg78372 Human sec
9	1054	42.5	1543	12 ADN60663	Adn60663 Human sec
10	972	39.2	2546	2 AAV44866	AAV44866 Clone CT5
11	972	39.2	2546	5 AAF98463	AAF98463 Human CDN
12	954	38.5	2510	5 AAF93372	AAF93372 Human CDN
13	891	35.9	2537	4 AAD08289	AAO08289 Human sec
14	861	34.7	2284	4 AAH18021	AAH18021 Human CDN
15	840	33.9	2547	4 AAD08315	AAO08315 Human sec
16	840	33.9	2660	6 ABQ54750	ABQ54750 Human ova
17	653	26.3	1114	4 AAH99794	AAH99794 Human pro
18	521	21.0	764	5 AAF93968	AAF93968 Primer sp
19	507	20.4	507	3 AAC01967	AAC01967 Human sec
20	506	20.4	787	4 AAH07526	AAH07526 Human CDN
21	500	20.2	500	6 ABV88332	ABV88332 Human col

22	430	17.3	433	6	ABL37128	ABL37128 Human col
23	402	16.2	558	5	AAF93612	AAF93612 Umbilical
24	359	14.5	387	4	ABA08397	ABA08397 Human sec
25	351	14.1	616	3	AAZ80248	AAZ80248 Human col
26	328	13.2	485	9	ACH24071	ACH24071 Human adu
27	322	13.0	349	2	AAV89737	AAV89737 EST clone
28	325	11.5	313	8	ABZ20328	ABZ20328 Group III
29	250	10.1	250	3	AAC01440	AAC01440 Human sec
30	236	9.5	245	3	AAC17015	AAC17015 Human sec
31	231	9.5	247	2	AAX40252	AAX40252 Human sec
32	231	9.3	281	3	AAA45221	AAA45221 Human sec
33	218	8.8	503	4	AAK91644	AAK91644 Human CDN
34	218	8.8	503	4	AAK93577	AAK93577 Human CDN
35	218	8.8	503	12	ADL30004	ADL30004 3' end of
36	218	8.8	503	12	ADL28071	ADL28071 5' end of
37	188	7.6	388	5	AAF66328	AAF66328 Novel hum
38	184	7.4	737	4	AAI96771	AAI96771 Human neu
39	179	7.2	474	5	AAF93633	AAF93633 Umbilical
40	156	6.3	393	4	ABA08547	ABA08547 Human sec
41	124	5.0	355	6	ABL37443	ABL37443 Human col
42	114	4.6	567	4	AAK92463	AAK92463 Human CDN
43	114	4.6	567	12	ADL28890	ADL28890 3' end of
44	105	4.2	483	9	ACH36684	ACH36684 Human end
45	65	2.6	2710	10	ADD94785	ADD94785 Mouse SIM

ALIGNMENTS

RESULT 1
ADD94783
ID ADD94783 standard; cDNA; 2481 BP.
XX
AC ADD94783;
XX

29-JAN-2004 (first entry)
XX

Human SIMP cDNA sequence.
XX

KW source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003054008-A2.
XX
PD 03-JUL-2003.
XX
PF 18-DEC-2002; 2002WO-CA001967.
XX
PR 20-DEC-2001; 2001US-00028384.
XX
PA (COMP-) COMPATIGENE INC.
XX
PI Perreault C, McBride K;
XX
DR WPI: 2003-559122/52.
XX
DR F-PSDB; ADD94784.

New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.

Claim 6; SEQ ID NO 1; 66pp; English.

This invention relates to a novel isolated or purified human protein.

Alignment Scores:	
Pred. No.:	3,956-214
Score:	2164.00
Length:	2356
Matches:	430
Percent Similarity:	71.39%
Best Local Similarity:	56.15%
Conservative:	114
Mismatches:	176


```
Db      2116 ATTGTGAGAAATTACAAAGTTAAAGATTTCGATAATGTTGGTAGAGATTTACATCAA--- 2172
QY      786 ArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLys 805
Db      2173 -----GCTACTGCTTTTGAAGATCA-----TCATCCGGCAGCTTCCAAAGAAAC 2217
QY      806 ArgGlyTyrIleLysAsnLysLeu 813
Db      2218 AGATCCATAAAGAGACCTAAATTG 2241
```

Search completed: December 15, 2004, 21:14:11
Job time : 1185.95 secs

1. The first part of the document is a letter from the

author to the reader, in which he explains the purpose of the

document and the reasons for its existence.

The second part of the document is a list of the

main points of the document, which are as follows:

1. The first point is that the document is a

list of the main points of the document, which are as follows:

1. The first point is that the document is a

list of the main points of the document, which are as follows:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:52:55 ; Search time 6394.33 seconds
(without alignments)
4707.175 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 4401

Sequence: 1 MAEPSAPSRKHSKSLNSPPW.....GYINKLVFKKGRKISKKTIV 826

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q/cn2_1/USPIO_spoos/US10028384/runat_14122004_131519_8030/app_query.fasta_1.3740
-DB=EST -QFMT=fastap -SUFFIX=ret -MINMATCH=0.1 -LOOPECU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORES=0 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10028384 @CGN 1.13960 @runat_14122004_131519_8030 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_ges1.*
9: gb_ges2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4263.5	96.9	2709	3 AK018758	Mus muscu
2	4216.5	95.8	2669	3 AK012153	Mus muscu
3	3918	4513	3	BC044321	Xenopus l
4	2763	62.8	2334	3 AY310157	Rattus no
5	2259.5	51.3	3802	3 AK030363	Mus muscu
6	2258.5	51.3	2656	3 BC028897	Mus muscu
7	2255.5	51.2	2730	3 AK081547	Mus muscu
8	2238	50.9	3643	3 AK077877	Mus muscu
9	2134	48.5	2118	9 AY418284	Homo sapi

10	2112	48.0	2727	3 AK087470	Mus muscu
11	2103.5	47.8	2103	9 AY418286	Mus muscu
12	2065.5	46.9	2419	3 CNS0ADY2	Arabidops
13	1691	38.4	2079	9 AY418285	Pan trogl
14	1563	35.5	950	5 BU515954	AGENCOURT
15	1513	34.4	2212	3 CR610915	full-leng
16	1465	33.3	852	7 CN458732	UI-M-HBO
17	1465	33.3	879	5 BQ947308	AGENCOURT
18	1391	31.6	823	7 CF736482	UI-M-HDO
19	1383	31.4	883	5 BU911227	AGENCOURT
20	1361	30.9	867	5 BU51585	603378796
21	1351	30.7	798	7 CN527592	UI-M-HQO
22	1337	30.4	1088	5 BX363211	BX363211
23	1333	30.3	766	7 CK633751	UI-M-HNO
24	1329.5	30.2	814	7 CN531243	UI-M-HNO
25	1324.5	30.1	896	6 CA980464	AGENCOURT
26	1318	29.9	957	5 BU554374	AGENCOURT
27	1313	29.8	825	6 CA318063	UI-M-FMO
28	1302.5	29.6	807	5 BU132963	60311925
29	1291	29.3	765	6 CA382416	NC
30	1289	29.3	735	4 BG698273	602660188
31	1289	29.3	754	7 CK636437	UI-M-HQO
32	1279	29.1	724	7 CF737921	UI-M-HDO
33	1276	29.0	877	6 CB208902	AGENCOURT
34	1270.5	28.9	776	7 CF730893	UI-M-GZO
35	1265.5	28.8	821	7 CF578043	AGENCOURT
36	1261	28.7	738	7 CNE27774	UI-M-HQO
37	1256.5	28.6	861	7 CF995035	AGENCOURT
38	1255	28.5	960	6 CD793779	EST665140
39	1252	28.4	759	4 B1662821	603286252
40	1249	28.4	724	5 BP701425	BP701425
41	1249	28.4	742	5 BQ443871	UI-M-EWO
42	1243.5	28.3	798	7 CK016163	AGENCOURT
43	1243	28.2	721	7 CNE29521	UI-M-HQO
44	1243.5	28.2	876	6 CD794446	EST65807
45	1240.5	28.2	779	5 BM963786	UI-M-EQO

ALIGNMENTS

RESULT 1
LOCUS AK018758 2709 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:130006C19 product:hypothetical oligosaccharyl transferase (OTase) STT3 subunit containing protein, full insert sequence.
ACCESSION AK018758
VERSION AK018758.2 GI:26384577
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitesunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED
11076861

4
THE RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 2709)
Arakawa, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Adachi, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Imorani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

On Dec 10, 2002 this sequence version replaced gi:12858635.
Please visit our web site (<http://genome.gsc.riken.jp/>) for further
details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5'-GAGAGAGAGCGGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGAGATCCAGAGATCAATTAATTAATTAACCCCGCCCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.

FEATURES

Location/Qualifiers
1..2709
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:1300006C19"
/db_xref="taxon:10090"
/clone="1300006C19"
/sex="male"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
76..2547

CDS

/note="unnamed protein product; hypothetical
Oligosaccharyl transferase (OTase) S7T3 subunit containing
protein (InterPro|IPR003674, evidence: InterPro)
Putative"
/codon_start=1

/protein_id="BAB1390.1"
/db_xref="GI:12858636"
/translation="MAEPSAPESKHKSLSNPSWGLMAGNSRHGHGPGTQASRA
AAPKGPAGLSGGSLGSPAGWQSLSTLFLAWLAGFSRLFAVIRFESIHEPDPW
FYNRSTHLAGHGFVEFLNPFDERAWPLGRIVGGTVYGLMITAGLIHMLNTLNT
VHIVDCVFLATPFGSLTSTFTLLRELNQGAGLLAACFIAIVPGYISRSVAGSFD
NEGIAIPALQPTVYLMVKSQTSFVFTMCCLSYFVMYSGMWGVIINLPLHFV
LLLMORYSKRYIAYSTLYIVGLLSMQIPVGFQPIRSEHMAAGVAFALQAYFL
QIQRDLTKQBFQTLFLGVSAAAGAVLSVITITGYIAPWSGFYSFSLWDITGYAKI
HIPIIASVSEHQPTTWVSFFDLVLTCPAGLWFCIKNINRDERVFAIYALISAVLY
AVVRLMLTLTPVYMLSAIAFNSVHYGLGDMKEENPVEDSDERKPNGLY
DKAGKRVHTEQKEPGLGNPKISIVTMLMLMFAVHCTWTWTNAYSPSYVL
ASYNHDGRNLDLDFREAYFWLQNTDEHARVMSWMDYQIAGMANRTLLVNNFTW
KSHIALVGKMSNETRAYKIMRSLDVIDYLVIFGVYIGYGGDDINKFLMWVIRABGE
HPKDLREGDYFTQGEPRVDKAGSPITLNCILMYKMSYRFGENQDLPFRPGFDTRN
AELGNKDIKFKHLEAETSEHLVRIYKVAFDNRETLGHKPRVTNIVPKQVLSKKT
TRKRGYVKNKLVFKGKTKSKIV"
2693..2698
/note="putative"
2709
/note="putative"

polyA_signal

polyA_site

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2709
Score: 4263.50 Matches: 805
Percent Similarity: 97.70% Conservative: 3
Best Local Similarity: 97.34% Mismatches: 14
Query Match: 96.88% Indels: 5
DB: 3 Gaps: 2

US-10-028-384-2 (1-826) x AK018758 (1-2709)

Qy 1 MetAlaGluProSerAlaProGluSerIshHisIysSerSerLeuAsnSerSerProTrp 20
Db 76 ATGGGGGAGCCCTCGGCCCGGAGCAAGCACAAGTCGTCCCTCACTGTCCTCCGTTG 135
Qy 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
Db 136 AGCGGCTTCATGGCTCTGGGGAACAGCGCCAGGCGCATGGCCCGGAGCCAGAGC 195
Qy 41 AlaHisIshAlaAlaGlyGlyAlaAlaProProIysProAla---ProAlaGlyLeuSer 59
Db 196 GCGTCCAGCGGCGCGCGC-----CCGAAGCGCGGCGCGCGCGCGCGCGTGTCC 243
Qy 60 GlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeu 79
Db 244 GGGGGCTTGTCCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 303
Qy 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIle 99
Db 304 GCGTGGCTGGCGGCTTCCAGCTCGCGCTCTTCGCCGTCATCCGCTTCGAGAGCATCATC 363
Qy 100 HisGluPheAspProTrpPheAsnTyArgSerThrHisIshIshIshIshIshIsh 119
Db 364 CACGAGTTCGACCGCTGTTTAACTATAGATCAACATCATCTTGCATCTCATGGATTC 423
Qy 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGly 139
Db 424 TATGAGTTCTTAATTTGTTTGTATGAAAGAGCATGGTACCCACTGGGGAAGATAGTGG 483
Qy 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
Db 484 GGCACCGTTTACCCAGCGGTTTCATGATCAACAGCTGGCCTTATTTCATTGATTTTAA 543
Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
Db 544 TTGAACATAACAGTTTCATAGAGATGTGTGTGTATTCCTTGCACCACTTTTAGCGGC 603
Qy 180 LeuThrSerIleSerPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199
Db 604 CTACATCATATCTACGTTCTCTGCTAACTAGAGAACTGTGGAACCAAGGAGCAGACTT 663

QY 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
 DB 664 CTAGCTGCTGCTTCATTGCTATCGTACAGGGTACATATCTCGTCAAGTGGCGGATCC 723
 QY 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLys 239
 DB 724 TTTGTAATGAAGGCAATGGCAATTTTGGCGTTCAGTTCACTTACTTATGGGTAAG 783
 QY 240 SerValLysThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMet 259
 DB 784 TCTGTGAAGACGGGCTCTGTCTTCTGGACAAATGTCTGCTCTGTTCATATTTCTCATG 843
 QY 260 ValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
 DB 844 GTCTCTGGTGGGAGGTTATGTGTTCAATCAACCTCACTCCCTCCCAAGTTTGTG 903
 QY 280 LeuLeuLeuMetGlnArgTyrSerIleArgValTyrIleAlaTyrSerThrPheTyrIle 299
 DB 904 TTGCTGCTGATGCGAGGTACAGCAAGAGAGTCTTACATAGCATATAGCACTTTGTACAT 963
 QY 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
 DB 964 GTGGGTTTAAATATATATCATCATGCAATACCTTTTGTGGATTTGAGCAATCAGAACAGC 1023
 QY 320 GluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
 DB 1024 GAGCACATGGCAGCTGCAGGTGTCTTTGCGTGTGCAAGCTTACGCTTTTGTGCAATAT 1083
 QY 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
 DB 1084 CTGAGAGACCGGTGACAAACAGAGATTCCAGACCCCTTTTCTTTTGGGTCTCACTA 1143
 QY 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
 DB 1144 GCTGAGCGCTGTGTCTTCTAGTGTCTATCTATCTGACATACAGGTTATATGACCA 1203
 QY 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399
 DB 1204 TGGAGTGCAGGTTTATCTACTATGGATCTGGTATCGGATGCAAAATATACATTCAT 1263
 QY 400 IleAlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHis 419
 DB 1264 ATTGCATCAGTGTCTGAAACATCAGCATCAGCATGGGTGTCTTCTTCTTGTATCTACAT 1323
 QY 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGlu 439
 DB 1324 ATTCTTGTATGTACTCTCCAGCAGGCTATGGTCTGCAATCAATAATCAACGATGA 1383
 QY 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
 DB 1384 AGAGTATTTGTGCTCTGTATGCGATCAGTGTCTGTACTTTGCGGAGTGTATGCGG 1443
 QY 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
 DB 1444 CTGATGCTGATCTTGACCCCGGTGTCTGATGTCTGTGCGCATCGCTTCTCAATGTT 1503
 QY 480 PheGluHisTyrLeuGlyAspAspMetLysArgGluAsnProValGluAspSerSer 499
 DB 1504 TTTGAGCACTATTTGGGGATGATGATAAGAGGAAACCCACCTGTGGAGGACAGCAT 1563
 QY 500 AspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLys 519
 DB 1564 GATGAGGATGACAAAGAAACCCAGGAAACCTTGTATGACAGGAGGATAAGTGAAGAG 1623
 QY 520 HisAlaThrGluGlnGluTyrGluGluGlyLeuGlyProAsnIleLysSerIleVal 539
 DB 1624 CATGTGCAGAGCAAGAAACCTGAAGAGGCTTGGGCGCCCAACATCAAAAGCATGTG 1683
 QY 540 ThrMetLeuMetLeuMetLeuMetPheAlaValHisCysThrTrpValThrSer 559
 DB 1684 ACCATGCTGATGCTCATGCTCTCTGATGATGTTTCCGGTTCACCTGCAGGTTGGTCAACAGC 1743
 QY 560 AsnAlaTyrSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsn 579

DB 1744 AACGCCCTACTCCAGTCCAGTGTGGTCTTGCCTCTCAATCATGATGGTACCAGGAAT 1803
 QY 580 IleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAla 599
 DB 1804 ATATTAGATGATTTTAGAGAGGCTACTTTTGGCTGAGACAAACACACGATGACACGCC 1863
 QY 600 ArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThr 619
 DB 1864 CGGTCATGTGCTGGTGGGACTACGCTATCAGATTTGCTGGATGGCCCAACAGGACCACT 1923
 QY 620 LeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSer 639
 DB 1924 CTGGTGTATCAACACACCTGGAACACAGCCACATCGCACTGGTGGAAAGACTATGCT 1983
 QY 640 SerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuVal 659
 DB 1984 TCCAATGAACCGGCCCTATAAAATCATGAGTCTCCCTTGTATGATGATGATGTTGGTT 2043
 QY 660 IlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetVal 679
 DB 2044 ATTTTCCGAGAGTGTATGGCTATTTCCGGGACGATATCAACAGTTCCTCTGGATGGTC 2103
 QY 680 ArgIleAlaGluGlyGlnHisProLysAspIleArgGluSerAspTyrPheThrProGln 699
 DB 2104 AGGATAGCTGAAGGGGAGCATCCCAAGACATCCCGGAAAGGTGACTATTTCCACCCAGCAG 2163
 QY 700 GlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLys 719
 DB 2164 GGAGAGTTCGAGTAGACAAAGCTGGGTCTCTCTACTCTGTTAACTGCTTATGATATAA 2223
 QY 720 MetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAsp 739
 DB 2224 ATGTCACTACTACAGATTGGAGAAATGCAGCTAGATTTTTCGCACTCCCGGCTTTGAC 2283
 QY 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAla 759
 DB 2284 CGAACACGTAATCTGAGATTGGAATTAAGACATTAATTAATCAAGCAATTTGGGAGAGCT 2343
 QY 760 PheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGlu 779
 DB 2344 TTTTACATCAGACGACTGGCTTGTGAGATATATAAAGTGAAGACCTGCACAAAGGGAG 2403
 QY 780 ThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLys 799
 DB 2404 ACATAGGTCAACAACTCGAGTCACCAACATCGTCCCAACACAGAGTATTTGTCAAG 2463
 QY 800 LysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLys 819
 DB 2464 AAGACTACTAAAGGAGCGTGGCTACGTAAATAAAGTAAAGTAGTGTAAAGAAAGGCAAG 2523
 QY 820 LysIleSerLysLysThrVal 826
 DB 2524 AAGACCTTAAGAGAGCTGTT 2544

RESULT 2

AK012153

LOCUS

DEFINITION

SOURCE

ACCESSION

VERSION

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

2669 bp mRNA linear HTC 03-APR-2004
 Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
 enriched library, clone:2610524N02 product:hypothetical
 Oligosaccharyl transferase (OTase) STT3 subunit containing protein,
 full insert sequence.
 AK012153
 AK012153.1 GI:12848721
 HTC; CAP trapper.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1
 Carninci P. and Hayashizaki Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 contact: amadan@systemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 94 Row: 1 Column: 19
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Similarity but not identity to protein
 This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
 1..4513
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:4684177"
 /tissue_type="Embryo, stage 31/32, xenopus"
 /clone_lib="NCHD_XGC_Emb4"
 /lab_host="DH10B"
 /note="vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4513
 Score: 3918.00 Matches: 751
 Percent Similarity: 93.38% Conservative: 25
 Best Local Similarity: 90.37% Mismatches: 39
 Query Match: 89.03% Indels: 17
 DB: 3 Gaps: 5

US-10-028-384-2 (1-826) x BC044321 (1-4513)

QY 176 ThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuThrPheAsnGln 135
 DB 781 GTATTAGTGGCTTACATCAATCTCCATTCTCTGCTCACTCGAAGTGTGGAACAG 840
 QY 196 GlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSer 215
 DB 841 GAGGCGGACCTTCTAGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 900
 QY 216 ValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 235
 DB 901 GTAGCTGGATCAATTTGACAATGAAGGCTCTCTTCTTCTTCTTCTTCTTCTTCT 960
 QY 236 LeuTyrValIleSerValIleThrGlySerValPheThrMetCysCysCysLeuSer 255
 DB 961 TTGTGGTAAATCTGTAACAGAGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 1020
 QY 256 TyrPheTyrMetValSerAlaTyrGlyTyrValPheIleIleAsnLeuLeuProLeu 275
 DB 1021 TATTTTATATGATTTCTGCTGCTGGGTGTTATGTTTATCATTAATCTTATTCAC 1080
 QY 276 HisValPheValLeuLeuMetGlnArgTyrSerIleValTyrIleAlaTyrSer 295
 DB 1081 CATGTGTTTGTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1140
 QY 296 ThrPheTyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPheGlnPro 315
 DB 1141 ACTTTTACATTTGCTGGTAAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1200
 QY 316 IleArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAla 335
 DB 1201 ATACGGACAAGTGAACATATGCGAGCTGCGAGGTCTTTGCTTCTTCTTCTTCT 1260
 QY 336 PheLeuGlnTyrLeuArgAspArgLeuThrIleGlnPheGlnThrIlePhePheLeu 355
 DB 1261 TTCTTCTAGTACTTGTAGAGCAAACTAAGTAAGCAGGATTTCCAGACACTGTTT 1320
 QY 356 GlyValSerLeuAlaAlaGlyValPheLeuSerValIleTyrLeuThrTyrThrGly 375
 DB 1321 GGTGTATCTTGGCAGCTGGAGCTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1380
 QY 376 TyrIleAlaProTyrSerGlyArgPheTyrSerLeuTyrPheThrGlyTyrAlaIle 395
 DB 1381 TACATTTCTCATGAGTGGAGATTTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1440
 QY 396 HisIleProIleIleAlaSerValSerGluHisGlnProThrTyrPheValSerPhe 415
 DB 1441 CATATCCCATCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 QY 416 PheAspLeuHisIleLeuValCysThrPheProIleGlyLeuThrPheCysIleLysAsn 435
 DB 1501 TTGTATCTGCATATATTAGTATGATTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1560
 QY 436 IleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGly 455
 DB 1561 ATCAACGATGAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 QY 456 ValMetValArgLeuMetLeuThrIleThrProValValCysMetLeuSerAlaIleAla 475
 DB 1621 GTAATGGTTCGTTGATGCTGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1680
 QY 476 PheSerAsnValPheGluHisTyrLeuGlyAspMetLysArgGluAsnProVal 495
 DB 1681 TTCTCCATGTTTTTGGACATTTATCTGGGTGATGATGATGAAGGGGAAACCC 1740
 QY 496 GluAspSerSerAspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGly 515
 DB 1741 GAAGACAGCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
 QY 516 LysValArgLysHisAlaThrGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 535
 DB 1795 AAAGTGAGAAAG 1854
 QY 536 LysSerIleValThrMetLeuMetLeuMetLeuMetPheAlaValHisCysThr 555

QY 226 AlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLysThrGlySer 245
 DB 469 GCGATTTCGCACTTCAGTTCACCTTACTTATGGGCCATTCT-----GGGGTT 519
 QY 246 ValPheTrpThrMetCys----- 251
 DB 520 GTCCTCTG-CCTTCTGCCCCATGCTCTCCAGGCCCTTGGTTCTCTCAGGGCATTCACT 578
 QY 252 -----CysCysLeuSer 255
 DB 579 CACCATTCTCTGTGGCCGCTGCTCTCTAGTGGGGCGCTGGTACATGTTGCTCGGCC 638
 QY 256 -TyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIleProLe 275
 DB 639 TGAGCTCTCAGGGTCTCTGCTGGGAGGTTATGTTTCATCATCAACCTCAICCTCT 698
 QY 275 uHisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSe 295
 DB 699 CCATGCTGTTGCTGCTCTGATGACAGAGGTACAGCAAGAGAGTCTACATA----- 750
 QY 295 rThrPheTyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnPr 315
 DB 750 ----- 750
 QY 315 oIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAl 335
 DB 751 -----GGTGTCTTTGGCTGCTGCGAGCTTATGC 779
 QY 335 aPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLe 355
 DB 780 GTTCTCGATCTCTGAGACCGGTTGACAAAGCAGAGGTTCCAGACCTTCTTCTTTT 839
 QY 355 uGlyValSerLeuAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGl 375
 DB 840 GGGTGTCTCACTAGCTGACGGCTGTGTTCTTGTGTCATCTATCTGACATAC----- 894
 QY 375 yTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIle 395
 DB 895 -----ACAGG-TATGCAAAAT 910
 QY 395 eHisIleProIleIleAlaSerValSerGluHisGlnProThrTrpValSerPhePh 415
 DB 911 ACATTTCCAAATATTTCATCTAGTCTGACATCAGCCCTACGACATGGGTGCTTCTT 970
 QY 415 ePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAs 435
 DB 971 CTTTGATCTACATATTCTTGTTGTTACCTTCCAGCAGGCCCTATGTTCTGCAICMAAAA 1030
 QY 435 nIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGl 455
 DB 1031 TATCAACGATCAAGAGATATT----- 1052
 QY 455 yValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAl 475
 DB 1052 ----- 1052
 QY 475 aPheSerAsnValPheGluHisTyrLeuGlyAspMetLysArgGluAsnProProVa 495
 DB 1052 ----- 1052
 QY 495 lGluAspSerSerAspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAla-G 515
 DB 1053 -----GGCAG 1057
 QY 515 lLysValArgLysHisAlaThrGluGlnLysThrGluGluGlyLeuGlyProAsnI 535
 DB 1058 GGAAGTGAGAAAGCATGTGACGAGCAGGAGGAGCGGAGGAGGGCTTGGGCCCAACA 1117
 QY 535 lLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysT 555
 DB 1118 TCAGAGCATCGTGACCATCTCTATGCTATGCTCTCTCATGATGTTGCTGTGCACTGCA 1177
 QY 555 hrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisA 575

DB 1178 CTGGGTCCAGCAGCAACCCCTACTCCAGCCCACTGGTCTCGCTCTCTCAATCATG 1237
 QY 575 spGlyThrArgAsnIleLeuAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnT 595
 DB 1238 ATGGAAACCAAGAAATATCTTAGATATTTAGAGAAGCATACTTTTGGTTGAGACAAAACA 1297
 QY 595 hrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetA 615
 DB 1298 CGGATGAACACGCCAGGCTCATCTCTGGTGGACTATGCTATCAGATTCTGGGATGG 1357
 QY 615 laAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValG 635
 DB 1358 CCACACAGCACTACTCTGGTAGCAATAACACTCGAACACACAGCCACATTCGCTGGTAG 1417
 QY 635 lLysAlaValSerSerAsnGlnThrAlaAlaTyrLysIleMetArgThrLeuAspValA 655
 DB 1418 GAAAAGCTATGTTCTTCTAATGAAACCGCCGCTATAAAATCATGAGTCTCTTGTATGTTG 1477
 QY 655 spTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysP 675
 DB 1478 ATTATGTCTTGTCTATTTTGGAGGAGTTATTGCTATTCTGTGTGATGATATCAACAAGT 1537
 QY 675 heLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspT 695
 DB 1538 TCTCTGTGATGTCAGGATAGCTGAAGGGGAGCAGCCCAAGACATTTCGGGAAGGGGACT 1597
 QY 695 yPheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnC 715
 DB 1598 ACTTCACCCAGCAGGAGGAGTTCCGAGTGGACAGAGCGGGTCTCTCTACTCTGTAAAT 1657
 QY 715 yLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrP 735
 DB 1658 GCCTTATGTATATAATGTCTACTACAGATTGTGAGAAATCGAGCTAGATTTTCGCACAC 1717
 QY 735 roProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysH 755
 DB 1718 CCCCCGGTTTTCACCGACACGTAATCTGAGATCGGAAATAGGACATTAAATTTCAAGC 1777
 QY 755 iLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaP 775
 DB 1778 ATTTGGAGAAAGCCCTTACATCAGACACTGGCTTGTGAGATATATAAAGTGAAGCAC 1837
 QY 775 roAspAsnArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnL 795
 DB 1838 CCGACACAGGAGGACATTAGTTCACAGCCTCGAGTCACCAACATCTTCCCCAACAAGA 1897
 QY 795 yTyrLeuSerLysLys-----ThrThrLysArgLysArgGlyTyrIleLysA 811
 DB 1898 AGTATTGTCAAGAAAGGATCTTGACATCCACAGAGAAAGAGGAGCACCACACCCCAAG 1957
 QY 811 snLysLeuValPhe-----LysLysGlyLysLysLysLys 822
 DB 1958 GCCTGACATGCTCAGTGAAGAGAGCAGAGAAAGGATGGAGACTTAGT 2007
 RESULT 5
 AK030363 3802 bp mRNA linear HTC 03-APR-2004
 LOCUS Mus musculus adult male pituitary gland cDNA, RIKEN full-length
 DEFINITION enriched library, clone:530402C17 product:OLIGOSACCHARYL
 TRANSFERASE STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
 homolog (Mus musculus), full insert sequence.
 ACCESSION AK030363
 VERSION AK030363.1 GI:26326354
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1
 AUTHORS Carninci P. and Havashizaki Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)


```

99279253 MEDLINE
10349636 PUBMED
2 REFERENCE
AUTHORS
Carrincci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374 PUBMED
11042159 REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
Suna, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwaga, K.,
Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL
Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE
20530913 PUBMED
11076861 REFERENCE
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
Nature 409, 685-690 (2001)
MEDLINE
20530913 PUBMED
11076861 REFERENCE
AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
MEDLINE
20530913 PUBMED
11076861 REFERENCE
AUTHORS
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Okano, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE
Direct Submission
JOURNAL
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genomic Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
FEATURES
source
1. 3802
/organism="Mus musculus"
/mol_type="cDNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:5330402C17"
/db_xref="taxon:10090"
/clone="5330402C17"
/sex="male"
/tissue_type="pituitary gland"

```

```

/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
133. .2250
/notes="unnamed protein product; OLIGOSACCHARYL TRANSFERASE
STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
homolog [Mus musculus] (SWISSPROT|P46978, evidence: FASTY,
99.7%ID, 100%length, match=2115)
putative"
/codon_start=1
/protein_id="BAC26921.1"
/db_xref="GI:26326355"
/translation="MTKLGFLRLSYEKQDTLLKLLSMAVLSTFTRPAVLRFESV
LHPEDPFTNRTTFLAEQGYKFNHFDRAWYPLGRIIGTITPGLMTAAIYHIV
LHFFDPTNRTTFLAEQGYKFNHFDRAWYPLGRIIGTITPGLMTAAIYHIV
SVAGSYNEGTAIFCMLLTYVMKIKVKTGSIYMAAKCALAFYVMSGGVYFLINL
IPLHLVLMLTGRFSRHYVAVCTVCLGTILSMQISFVGFQFVLSSEMAAGVFGGL
COHAFVDVLCSLKPNQPEVLSVLSGVFLVTVGALLMTGKISPTGRFYSLL
DPSAKNNIILIASVSEHQPTTASSYFYLQLLVFNFPVGLYCCSNLSLDAIFILMY
DTSWTFSAVMVRLMLVAFVCMILSGIGVQVSTVMKNLDISAPDKKSKQQDSTY
PIKNEVASGMILVNAFLITYTTPHSTWVTSYSSPISVLSARGDGSRIIFDPEEA
YYLRLHNTPEDAKVMWDYQITAMANTILVDNNTWNTHTSRVGGQAMASTEKA
YEIMRELDVSVLVIFGLTGYSDDDINKPLVMVRIGGSTETGRHIKENDYTPGEGF
RVDRGSEVLLNCLMYKVCYRFQGVVTEAKRPPGDFRVRNAEIGNKDFELVLEAY
TTEHVLRYLYKVDLNRGLSRT"
ORIGIN
Alignment Scores:
Pred. No.: 7,41e-214 Length: 3802
Score: 2259.50 Matches: 435
Percent Similarity: 70.46% Conservative: 104
Best Local Similarity: 56.86% Mismatches: 155
Query Match: 51.34% Indels: 71
DB: 3 Gaps: 8
US-10-028-384-2 (1-826) x AK030363 (1-3802)
QY 58 LeuSerGlyLeuSerGlnProAla---GlyTrpGln----- 69
Db 19 CTGCTGTGAGGTTGAGTGCGCGCTGAACGGATGGCAGGGGAGCAGAGTGGGTTCTG 78
QY 69 ----- 69
Db 79 AGGAGCATCGGTGAGGTATTGTAATATCATCATCTGCCACCCATTGATGTCAGATGACT 138
QY 70 -----SerLeuLeuSerPheThrIle 76
Db 139 AAGCTTGAGTTTGGCGATTGCTATGAGAGCAGCAGACACTTCTAAAGCTTCTCATC 198
QY 77 LeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGlu 96
Db 199 CTGTCGATGGCTGCTGTGTATCTTTTCTACTCGCTCTTTTCTGTGTGTGAGATTGAA 258
QY 97 SerIleIleHisGluPheAspProTrpPheAsnTrpArgSerThrHisHisLeuAlaSer 116
Db 259 AGTGTCATCCATGAGTTTGTATCGGTACTTTTAATATCGGACTACCCGGTTTCTGGCTGAG 318
QY 117 HisGlyPheTrpGluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProGluArg 136
Db 319 GAGGGTTTATAAATTCATATCCATATCGTTTGTAGACCGGCTGGTACCCCTTTGGCCGA 378
QY 137 IleValGlyGlyThrValTrpGlyLeuValIleThrAlaGlyLeuIleHisTrpIle 156
Db 379 ATCATTTGAGGAGCAATTTACCAGGTTTAAATGATCATTCTGCTGCAATCTACCATGTA 438
QY 157 LeuAsnThrLeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThr 176
Db 439 CTCCATTTCTCCATATCACCATTGACATTCGGAATGTCTGTGTTTCTTGGCCCGCTT 498
QY 177 PheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGly 196
Db 499 TTCTCTCTTTTCCACCACCATCTGTACGTACCATCTTACCACAGAGCTCAAGGATGACGGA 558
QY 197 AlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTrpIleSerArgSerVal 216

```

[illegible]

1558	ACTTGGGTGACAGTGAAGCCATTCTTCTCCCTCCATTTGTTCTGTCTGCTCGTGGG	1617
QY	AspGlyThrArgAsnIleLeuAspPheArgGluAlaTyrPheTrpLeuArgGlnAsn	594
1618	GATCGCATAGGATCATTTTTCATGACTCCGAGAAGCGTATTATTGGCTCCGTCACAAT	1677
QY	ThrAspGluHisIslaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMet	614
1678	ACTCCAGGAGATCAAAAGTCATGTCATGGTGGGATTATGCTACCAATATTCTCGAATG	1737
QY	AlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleLeuVal	634
1738	GCAAATCGGACAAATTTTTCATGACATAAACAATGGATTAATATCCCATATTTCTCGAGTA	1797
QY	GlyIleAlaMetSerSerAsnGlnThrAlaAlaTyrIleValMetArgThrLeuAspVal	654
1798	GGCAGCGCATGTCATCCACAGAGAAAAGCCATATGNAATCATGAGGGAGCTTGATGTC	1857
QY	AspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLys	674
1858	AGCTATATGCTTCTCATTTTTCGAGCGCCTTACTCGGTATTTCTCGGATGATATCAACAAG	1917
QY	PheLeuTrpMetValArgIle--AlaGluGlyGluHisProLysAspIleAsnGluSer	693
1918	TTTCTTTGGATGTCGGATTGGAGGAGCAGACAGAGACACATATAGGAGAAAT	1977
QY	AspTyrPheThrProGlnGlyGluPheArgValAspIleAsnGlySerProThrLeuLeu	713
1978	GACTACTATATCTCTACTCGGAAATCCGTTGATCGTGAGGGCTTCTCCGGTCTGCTC	2037
QY	AsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArg	733
2038	AACCTGCCTTATGACAAATGTGTTACTACCGCTTGGCAGGTCTACACAGAGCCAAG	2097
QY	ThrProGlyPheAspArgThrArgAsnIleGluIleGlyAsnLysAspIleLysPhe	753
2098	CGTCCACAGCGCTTTCACCGCTGTCGAAATGCTGAGATTGTAATAAGACTTTTGAGCTT	2157
QY	LysHisLeuGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLys	773
2158	GATCTCTCGGAGGACGGTATACCACAGACACACTGGCTAGTCAGGATATACAGGTAAG	2217
QY	AlaProAspAsnArg 778	
2218	GACCTGGATAATCGA 2232	
RESULT 6	2656 bp mRNA linear HTC 04-AUG-2004	
BC028897	Mus musculus cDNA clone IMAGE:4168013, containing frame-shift	
LOCUS	errors.	
DEFINITION	BC028897.1 GI:22137711	
ACCESSION	BC028897	
VERSION	BC028897.1	
KEYWORDS	HTC.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
	1 (bases 1 to 2656)	
	Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.	
	Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,	
	Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,	
	Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,	
	Diatchenko, L., Marushina, K., Farmer, A.A., Rubin, G.M., Hong, L.,	
	Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavani, T.L.,	
	Sheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,	
	Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,	
	Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J.,	
	McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,	
	Worley, K.C., Hale, S.G., Garcia, A.M., Gay, L.J., Hulyk, S.W.,	
	Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,	
	Fahey, J., Halton, E., Kettman, M., Madan, A.C., Rodriguez, S.,	
	Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,	

/tissue type="head"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="16 days embryo"
 125..2243

misc_feature

/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
 (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
 (SWISSPROT P46978, evidence: FASTV, 99.7%ID, 100%length,
 match=2115)
 putative"
 2717..2722

polya_signal

/note="putative"
 2730

/note="putative"
 2730

ORIGIN

Alignment Scores:

Pred. No.: 1,1e-213 Length: 2730
 Score: 2255.50 Matches: 436
 Percent Similarity: 70.50% Conservative: 104
 Best Local Similarity: 56.92% Mismatches: 154
 Query Match: 3 Indels: 72
 DB: 3 Gaps: 8

US-10-028-384-2 (1-826) x AK081547 (1-2730)

QY 58 LeuSerGlyGlyLeuSerGlnProAla---GlyTrpGln----- 69
 DB 11 CTGCTGTCAAGGTTGAGTGGCCCGCTGAACGGATGGCAGGGGAGCAGAGTGGTTCCTG 70
 QY 69 ----- 69
 DB 71 AGGAGCATCCGTGAGGTATTTGAATATCATCATGTTGCCACCCCAATTGATCAAGATGACT 130
 QY 70 -----SerLeuLeuSerPheThrIle 76
 DB 131 AGCTTGGATTTTCGGATTGTCCTATGAGAGCAGCAGACACATCTTCAAGCTTCTCATC 190
 QY 77 LeuPheLeuAlaTrpLeuAlaGlyPheSerArgLeuPheAlaValIleArgPheGlu 96
 DB 191 CTGCGATGGCTGCTGCTGCTATCTTTTCTACTGCTCTTTTGGCTGCTGAGATTGAA 250
 QY 97 SerIleIleHisGluPheAspProTrpPheAsnTrpArgSerThrHisHisLeuAlaSer 116
 DB 251 AGTGTCATCCATGATTGATCCGACCTTAATATATCGGACTACCCCGTTTCTGGCTGAG 310
 QY 117 HisGlyPheTrpGluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArg 136
 DB 311 GAGGGGTTTATAAAATTCATAACTGGTTTGTGATGACCGGGCTTGGTACCCCTTTGGGCGGA 370
 QY 137 IleValGlyGlyThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIle 156
 DB 371 ATCATTTGAGGAGAACAAATTATCCAGGTTTAAATGATCACTTCGTGCAATCTACCATGTA 430
 QY 157 LeuAsnThrLeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThr 176
 DB 431 CTCCATTTCTTCATATCACCATTGACATTCGGAATGCTGTGTTTCTCTGGCCCGCTT 490
 QY 177 PheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGly 196
 DB 491 TTCTCTCTTTCCACCCATCGTTACGPAACACCTTACCAAGAGCTCAAGAGTACGAGGA 550
 QY 197 AlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTrpIleSerArgSerVal 216
 DB 551 GCTGGGCTTCTGCTGCTGCCATGATTCGTAGTTCCTGGGTATATTTCTCGATCTGTA 610
 QY 217 AlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpLeu 236
 DB 611 GCTGCTCTATGATGATGATGAGAAATTCATATCTTTTGCATGCTGCTTACTACTACAT 670
 QY 237 TrpValLysSerValIleThrGlySerValPheTrpThrMetCysCysLeuSerTrp 256
 DB 671 TGGATCAAGGCAGTCAAGACTGGTTCCATCTATTGGGCTGCCAAGTGTGCCCTTCGCTTAT 730

QY 257 PheTrpMetValSerAlaTrpGlyGlyTrpValPheIleLeuLeuLeuProLeuHis 276
 DB 731 TTCATATGGTCTCTTATGAGGGAGGCTATGTGTCTGATCACTTGTATCTCTACAT 790
 QY 277 ValPheValLeuLeuLeuMetGlnArgTrpSerLysArgValTrpIleAlaTrpSerThr 296
 DB 791 GTCTCTGGTGTAAATGCTCAGCAGGCGGTTTCTCACCGGATCTACGTAGCTACTGACT 850
 QY 297 PheTrpIleValGlyLeuLeuSerMetGlnIleProPheValGlyGlnProIle 316
 DB 851 GTTACTGCTGGGACCAATCTTCTATGAGATTTCCTTGTGGTTTCCAGCCCGTC 910
 QY 317 ArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPhe 336
 DB 911 CTTTCATCAGAACACATGGCAGCCCTTTGGAGTGTGTTGGTCTCTGTCAGATCCATGCTTC 970
 QY 337 LeuGlnTrpLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGly 356
 DB 971 GTAGATTACCTGGCAGCAAGTTGAATCCACAGCAATTCGAGTTCTTTCCGAGTGT 1030
 QY 357 ValSerLeuAlaAlaGlyAlaValPheLeuSerVal---IleTrpLeuThrTrpValSerPhe 375
 DB 1031 ATCTCCCTG---GTTGGCTTGTCTCTCTCACTGTGGAGCTCTCTCATGCTTAACAGGA 1087
 QY 376 TyrIleAlaProTrpSer-GlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaLysIle 395
 DB 1088 AAATTTCTCCCTCGACAGGGCGTTTCTACTCTCTGCTGATCCCTCTTATGCTAAGAA 1147
 QY 395 eHisIleProIleLeuAlaSerValSerGluHisGlnProThrTrpValSerPhePhe 415
 DB 1148 TAACATTTCCCATTTATTCATCTGTTTCCAGTGGCTCTATATCTGCTTCTCTACTA 1207
 QY 415 ePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAs 435
 DB 1208 TTTTGATCTACAGCTCTTCTTCATGTTTCCAGTGGCTCTATATCTGCTTCTTACCAA 1267
 QY 435 nIleAsnAspGluArgValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAlaG 455
 DB 1268 CTTCTCTGATCTCGGATTTTATCATCATGATGTTGTTGACCAACCACTGTTCTTTCAGC 1327
 QY 455 ValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAl 475
 DB 1328 TGTAAATGCTGCTTAATGCTGTTATGTCACCTGTTATGTCATCTTCTGCGATTGG 1387
 QY 475 aPheSerAsnValPheGluHisTrpLeuGlyAsp---AspMetLysArgGluAsnProPr 494
 DB 1388 TGTTCCTCCAGGTGCTGCCACATATGAAAAATCTGGACATAAGTCGCCCA----- 1439
 QY 494 oValGluAspSerSerAspGluAspLysArgAsnGlnGlyAsnLeuTrpAspLysAl 514
 DB 1440 -----GACAGAA 1447
 QY 514 aGlyLysValArgLysHisAlaThrGluGlnGlyLysThrGluGluGlyLeuGlyProAs 534
 DB 1448 GAGCAAGAAAGCAACAGGATCTTACTTACCTTATTAAGATGAG----- 1490
 QY 534 nIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCys 554
 DB 1491 -GTGCGAGTGGGATGATGATGTCATGGCTTCTTCTCATCCTACCTACGTTCTTCTTC 1549
 QY 554 sThrTrpValThrSerAsnAlaTrpSerProSerValValLeuAlaSerTrpAsnHis 574
 DB 1550 GACTTGGGTGACCAAGTGAAGCTTATCTTCTCCCTCCATTTGTTCTGCTGCTGCTGGTGG 1609
 QY 574 sAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTrpPheTrpLeuArgGlnAs 594
 DB 1610 GGATGGCAGTAGGATCAATTTTGAAGCTTCCGAGAGCGTATATTGGCTCCGTCACAA 1669
 QY 594 nThrAspGluHisAlaArgValMetSerTrpTrpAspTrpGlyTrpGlnIleAlaGlyMe 614
 DB 1670 TACTCCAGAGGATGCAAAAGTCATGTATGCTGGGATATTGCTGCTACCAATTAAGTCAAT 1729
 QY 614 tAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuVa 634

```

Db      1730  GCGAAATCGGCAATTTTGTGGACATAACACATGCAATATACCATATTCCTCGAGT 1789
Qy      634  IGLYLSAlaMetSerSerAsnGluThrAlaLalTyLysIleMetArgThrLeuAspVa 654
Db      1790  AGGCGAGCAATGCGATCCACAGAGAAAAGCCTATGAATCATGAGGAGCTTGATGT 1849
Qy      654  lAspTyValLeuValIlePheGlyGlyValIleGlyTySerGlyAspAspIleAsnIly 674
Db      1850  CAGCTATGTGCTGCTATTTTGGAGCCCTTACTGGTATTCCTCGATGATATCAACAA 1909
Qy      674  sPheLeuTrpMetValArgile---AlaGluGlyGluHisProLysAspIleArgGluLe 693
Db      1910  GTTCTTGGATGCTCGGATGGAGGAACACAGACAGAGAGACACATTAAAGAGAA 1969
Qy      693  rAspTyPheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLe 713
Db      1970  TGACTACTATACTCTCTACTGGGAATTCCTGTGTGATCGTGGAGGTCTCCGGTCTGCT 2029
Qy      713  uAsnCysLeuMetTyLysMetSerTyTyArgPheGlyGluMetGlnLeuAspPheAr 733
Db      2030  CAACCTGCCTTATGTAACAAATGTTACTACCGCTTTGGCAGGCTACACAGAACCCAA 2089
Qy      733  gThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleAsp 753
Db      2090  CGCTCCACCCAGGCTTTGACCGGTGTCGAAATGCTGAGATGGTAATAAGACTTTGAGCT 2149
Qy      753  eLysHisLeuGluAlaPheThrSerGluHisTrpLeuValArgIleTyLysVally 773
Db      2150  TGATGTCTCTGGAGGAGCGTATACACAGAACTGGTGTAGTCAGGATATACAGGTAAA 2209
Qy      773  sAlaProAspAsnArg 778
Db      2210  GGACTGTGATAATCGA 2225

RESULT 8
LOCUS   AK077877
DEFINITION Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
          enriched library, clone:6030401L06 product:OLIGOSACCHARYL
          TRANSFERASE STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
          homolog [Mus musculus], full insert sequence.
ACCESSION AK077877
VERSION   AK077877.1 GI:26097543
KEYWORDS  HTC; CAP trapper.
SOURCE    Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1
          Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
          Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
          Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new genes
          Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE 2
          Carninci,P. and Hayashizaki,Y.
          High-efficiency full-length cDNA cloning
          Meth. Enzymol. 303, 19-44 (1999)
          99279253
          10349636
REFERENCE 3
          Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
          Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
          Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new genes
          Genome Res. 10 (10), 1617-1630 (2000)
          20499374
          11042159
REFERENCE 4
          Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
          Konno,H., Akiyama,Y., Nishi,K., Kutsuna,T., Tashiro,H., Itoh,M.,
          Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
          Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasaiwagi,K.,
          Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wachiwagi,K.,
          Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
          Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
          RIKEN integrated sequence analysis (RISA) system-384-format
          sequencing pipeline with 384 multicapillary sequencer

```

```

JOURNAL  Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE  20530913
PUBMED   11076861
REFERENCE 4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
          PANTOM Consortium.
          Functional annotation of a full-length mouse cDNA collection
          Nature 409, 685-690 (2001)
JOURNAL  5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
          Group Phase I & II Team.
          Analysis of the mouse transcriptome based on functional annotation
          of 60,770 full-length cDNAs
          Nature 420, 563-573 (2002)
JOURNAL  6
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
          Fukuda,S., Furuno,M., Hasegaki,T., Hara,A., Hashizume,M.,
          Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
          Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
          Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
          Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
          Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
          Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
          Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
          Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akaira,S.,
          Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
          Muramatsu,M. and Hayashizaki,Y.
          Direct Submission
          Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
          Physical and Chemical Research (RIKEN), Laboratory for Genome
          Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
          RIKEN Yokohama Institute, 1-7-22 Sushiro-cho, Tsurumi-Ku, Yokohama,
          Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
          URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
          Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in RIKEN.
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues.
          Please visit our web site for further details.
          URL:http://genome.gsc.riken.jp/
          URL:http://fantom.gsc.riken.jp/
FEATURES   Location/Qualifiers
            1..3643
               /organism="Mus musculus"
               /mol_type="mRNA"
               /strain="C57BL/6J"
               /db_xref="PANTOM,DB:6030401L06"
               /db_xref="taxon:10090"
               /clone="6030401L06"
               /sex="male"
               /tissue_type="testis"
               /clone_lib="RIKEN full-length enriched mouse cDNA library"
               /dev_stage="13 days embryo"
            93..2208
               /note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
               (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
               (SWISSPROT[P46978], evidence: FASTA, 99.7%ID, 100%length,
               match=2115)
               putative"
misc_feature
            93..2208
               /note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
               (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
               (SWISSPROT[P46978], evidence: FASTA, 99.7%ID, 100%length,
               match=2115)
               putative"
ORIGIN
Alignment Scores:
Pred. No.: 9,796-212 Length: 3643
Score: 2238.00 Matches: 427
Percent Similarity: 74.58% Conservative: 104
Best Local Similarity: 59.97% Mismatches: 151
Query Match: 50.85% Indels: 31
DB: 3 Gaps: 6
US-10-028-384-2 (1-826) x AK077877 (1-3643)

```


LOCUS AY418284 2118 bp DNA linear GSS 17-DEC-2003
 DEFINITION Homo sapiens ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.
 ACCESSION AY418284
 VERSION AY418284.1 GI:39774244
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2118)
 Clark, A.G., Gianowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J., Adams, M.D., and Cargill, M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering them based on alignment.
 FEATURES
 source
 1. 2118
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 <1..>2118
 /gene="ITM1"
 /locus_tag="HCM6506"
 gene
 Alignment Scores:
 Pred. No.: 1.04e-201 Length: 2118
 Score: 2134.00 Matches: 405
 Percent Similarity: 73.53% Conservative: 95
 Best Local Similarity: 59.56% Mismatches: 144
 Query Match: 48.49% Indels: 36
 DB: 9 Gaps: 7
 US-10-028-384-2 (1-826) x AY418284 (1-2118)
 QY 105 TrpPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsn 124
 DB 151 TACTTTAATTAATTCGACCTACAGGCTTCTGGCTGAGAGGGGTTTATTAATTCATAC 210
 QY 125 TrpPheAspGluArgAlaTyrTrpProLeuGlyArgIleValGlyGlyThrValTyrPro 144
 DB 211 TGGTTTGTATGACCGAGCGTGTACCTTTGGGACGAATCATTTGGAGGAACAAATTTACCA 270
 QY 145 GlyLeuMetIleThrAlaGlyLeuIleHisTyrPheLeuAsnThrLeuAsnIleThrVal 164
 DB 271 GGTATTATGATCACTCTGTGTCGCAATCATGATGACTCCATTTTCCACATCCATC 330
 QY 165 HisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSer 184
 DB 331 GACATTCGAAATGTGTGTGTTCTCGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 390
 QY 185 ThrPheLeuLeuThrArgGluLeuTyrAsnGlnGlyValGlyLeuLeuAlaIleCysPhe 204
 DB 391 ACGTACCACTTACCAAGAGCTCAAGATGCGAGGGCTGGGCTCTCTCTCTCTCTCTCTCTCT 450
 QY 205 IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAsnGlnGly 224
 DB 451 ATTGCTGTAGTTCCTGTGATATCTCCCGATCTGTGGCTGGCTCTCTCTCTCTCTCTCTCTCT 510

QY 225 IleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLysThrGly 244
 DB 511 ATGGCAATTTTGTGATGCTACTACCTACTACTATGATGGATCAAGGAGGTAAGAGCTGGT 570
 QY 245 SerValPheTyrThrMetCysCysCysLeuSerTyrPheTyrMetValSerAlaTyrGly 264
 DB 571 TCCATCTGTGGGAGCTAAAGTGTGCCCTTCTATTCTTACATGGTCTCGTCATCGGGA 630
 QY 265 GlyTyrValPheIleAlaLeuLeuIleProLeuHisValPheValLeuLeuLeuMetGln 284
 DB 631 GGTATGTTGTTCTGATCACTTAATCTCTCCAGCTCTCTGCTGCTGATGCTCACAGGC 690
 QY 285 ArgTyrSerLeuArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuLeuLeu 304
 DB 691 CGTTTCTCTCACCGATCTATGTGGCTACTGTACTGTCTTACTGCTGGGACATATACTT 750
 QY 305 SerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAla 324
 DB 751 TCTATGCAATCTCTTGTGGTTCAGGCTGCTCTTTCATCAGAGCACATGGCAGCC 810
 QY 325 AlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeu 344
 DB 811 TTGGGGTCTTTGGTCTCTGCCAGATCCATGCTTGTGGATTACCTGCGCAGCAAGTTG 870
 QY 345 ThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaVal 364
 DB 871 AATCCACACATTTGAAGTTCTTTTCGGAGGCTCATCTCTCTG---GTAGGCTTTGTC 927
 QY 365 PheLeuSerVal---IleTyrLeuThrTyrThrGlyTyrIleAlaProTyrSerGlyArg 383
 DB 928 CTCTCAGCGTGGGAGCTCTCTCTNNNNNNNNNAAAAATATCTCTCCCTGGAGGGCGCT 987
 QY 384 PheTyrSerLeuTyrAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerVal 403
 DB 988 TTCTACTGCTGCTGATCCCTCTTATGCTAAGAACACATCCCATCATTTGCTTCTGTG 1047
 QY 404 SerGluHisGlnProThrThrTyrValSerPhePheAspLeuHisIleLeuValCys 423
 DB 1048 TCTGAGCATCAGCCACACCTGCTCTCATCTATTGACCTGCGAGCTCTCTGCTTTC 1107
 QY 424 ThrPheProAlaGlyLeuTyrPheCysIleLysAsnIleAsnAspGluArgValPheVal 443
 DB 1108 ATGTTTCCAGTTGGCTCTTATTACTGCTTTAGCAACCTGCTGATGCCGGAATTTATC 1167
 QY 444 AlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThr 463
 DB 1168 ATCATGTATGTTGACCGAGCATGTACTTTTCACTGCTGATGCTGCTGCTGCTGCTGCT 1227
 QY 464 LeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyr 483
 DB 1228 TTGGCACTGTTATGTGCAATCTCTCTGCACTGGAGTCTCCAGGTGCTGTCCACATAC 1287
 QY 484 LeuGlyAsp---AspMetLysArgGluAsnProProValGluAspSerSerAspGluAsp 502
 DB 1288 ATGAAGAATCTCGACATCAAGTCGTCCA-----GACAAAGAGCAGCAAGCAAGATTCCACC 1314
 QY 503 AspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThr 522
 DB 1315 -----GACAAAGAGCAGCAAGCAAGCAAGATTCCACC 1347
 QY 523 GluGlnGlyThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeu 542
 DB 1348 TACCTTATTAAAGTAA-----GTGGCAATGGGATGATGATCTGCTC 1389
 QY 543 MetLeuMetLeuMetMetPheAlaValHisCysThrTyrTrpValThrSerAsnAlaTyr 562
 DB 1390 ATGGCTTTCTTCTCATCACCCTACACCTTCTTCACTGGGTGACCAAGGAGGCTAC 1449
 QY 563 SerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAsp 582
 DB 1450 TCTCTCCGTCATGATGATCTATCTCCCGTGGGTGGGATGATGATGATGATGATGATGAT 1509
 QY 583 AspPheArgGluAlaTyrPheTyrLeuArgGlnAsnThrAspGluHisAlaArgValMet 602


```

Db      1510 GACITCCGAGAGCATATATGGCTTCGTATATCTCCAGAGGATGCGAAGTCATG 1569
QY      603 SerTTPtPAspTyrGlyTyrGlnIleAlaGlyMetAlaSerArgThrThrLeuValAsp 622
Db      1570 TCTTGGTGGATTATGGCTATCAGATTACAGCTATGCGAACCAGCAATTTAGTGGAC 1629
QY      623 AnAnThrTTPAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGlu 642
Db      1630 AATAACACATGGAATTAATCCATATTTCTCAGTAGGCGAGCAATGGCGTCCACAGAG 1689
QY      643 ThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGly 662
Db      1690 GAAAGAGCTATGACATCATGAGGAGCTCGATCGACGTATGCTGGTCATTTTGG 1749
QY      663 GlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetValArgIleAla 682
Db      1750 GGCCTCACGTGGTATTCCTCTGATGATATCAACAGTTCTTGGATGGTCCGATGGA 1809
QY      683 -----GluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrPro 698
Db      1810 GGGAGCACAGATACAGGCAACAT-----ATCAAGGAGATGACTATTATCTCCA 1860
QY      699 GlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyr 718
Db      1861 ACTGGGGAGTCCGGTGGACCGGAGGTTCTCCAGTGTCTCAACTGCCCTCATGTAC 1920
QY      719 LysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPhe 738
Db      1921 AAGATGTGTACTATCGTCTGGACAGGTTTACACAGANNNNAGCGTCTCCAGGCTTT 1980
QY      739 AspArgThrArgAsnAlaGlnIleGlyAsnLysAspIleLysPheLysHisLeuGlu 758
Db      1981 GACCGTGTCCGAAAGCTGAGATGGGAATAAAGACTTTGAGTTGATGCTCCGGAGGA 2040
QY      759 AlaPheThrSerGluHisThrPleuValArgIleTyrLysValLysAlaProAspAsnArg 778
Db      2041 GCATATACACAGACAATTGCTGCTCAGGATATACAAAGGTAAAGGACCTGGATATCGA 2100

RESULT 10
AK087470
LOCUS
DEFINITION
Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus
musculus], full insert sequence.
ACCESSION
AK087470
VERSION
AK087470.1
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Chordata; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
2 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
3 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
4 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
5 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
6 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
7 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
8 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
9 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
10 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

```

```

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530313
PUBMED
11076861
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2727)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hata, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasekawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
location/Qualifiers
1. 2727
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:E130302E03"
/db_xref="taxon:10090"
/clone="E130302E03"
/tissue_type="eyeball"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 Day neonate"
108. 2233
/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
(B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
(SWISSPROT[P46978], evidence: FASTA, 99.7%ID, 100%length,
match=2115)
Putative"
2708. 2713
/note="putative"
2727
/note="putative"

misc_feature
108. 2233
/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
(B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
(SWISSPROT[P46978], evidence: FASTA, 99.7%ID, 100%length,
match=2115)
Putative"
2708. 2713
/note="putative"
2727
/note="putative"

polyA_signal
2727
/note="putative"

polyA_site
2727
/note="putative"

ORIGIN
Alignment Scores:
2.47e-199 Length: 2727
Score: 2112.00 Matches: 419

```

```

Percent Similarity: 68.49%      Conservative: 107
Best Local Similarity: 54.56%    Mismatches: 164
Query Match: 47.99%             Indels: 79
DB: 3                            Gaps: 8

US-10-028-384-2 (1-826) x AK087470 (1-2727)

Qy 61 GlyLeuSerGlnProAla---GlyTrpGln----- 69
Db 3 GGGTTGAGTGCCCGCTGACCGGATGGCAGGGGAGCAGAGTGGGTCTCTGAGGAGCATC 62
Qy 69 ----- 69
Db 63 CGTGAGGTATTGGAATATCATCATAGTTGCCACCCCATGATGTCAAGATGACTAAGCTTGGGA 122
Qy 70 -----SerLeuLeuSerPheThrIleLeuPheLeu 79
Db 123 TTTTTCGGATTGCTATGAGAGCAGGACACACTTCTAAAGCTTCTCATCTCTGCGATG 182
Qy 80 AlaTrpLeuAlaGlyPheSerArgLeuPheAlaValIleArgPheLeuSerIleIle 99
Db 183 GCTCCTGTATCTTTTCTACTCTCTCTTTTGTCTGCTGAGATTTGAAAGTGTCTATC 242
Qy 100 HisGluPheAspProTrpPheAsnTrpArgSerThrHisHisLeuAlaSerHisGlyPhe 119
Db 243 CATGAGTTTGTATCGTACTTTAATTATCGGACTACCCGGTTCTGGCTCAGGAGGGGTTT 302
Qy 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGly 139
Db 303 TATAAATTCCTAATCACTGGTTTGTATGACCGGGCTTGTACCTTTTGGCGCGATGATTGGA 362
Qy 140 GlyThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
Db 363 GGAACAATTACCCAGGTTTAATGATCACTTCTGCTGCACTACCATGCTACTCCATTTC 422
Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
Db 423 TTCATATCACTACCATGATGATTCGGAATGTCTGTGTCTTCTCGCCCGCTTTCTCCCTCT 482
Qy 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValGlyLeu 199
Db 483 TTCACCACTCGTTAGTACCACTTACCAAGAGCTCAAGATGCGAGGCTGGGTT 542
Qy 200 LeuAlaAlaCysPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySer 219
Db 543 CTTGGTGGCGCATGATGTCTAGTCTCTGGGATATTTCTCGATCTGTAGTGGCTCC 602
Qy 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpLeuTrpVallys 239
Db 603 TATGATAATGAAGGAATTGTACTTTTGCATGCTGCTTACTACTACATGTGGATCAAG 662
Qy 240 SerVallysThrGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMet 259
Db 663 GCAGTGAAGACTGGTTCCTATATGGGCTGCCAAGTGTGCCCTCGCTTATTTCTACATG 722
Qy 260 ValSerAlaTrpGlyGlyTrpValPheIleIleAsnLeuIleProLeuHisValPheVal 279
Db 723 GTCTCTTCATGGGAGGCTAATGTCTTCTGATCACTTGTATCTCTACATGTCTCTGCTG 782
Qy 280 LeuLeuLeuMetGlnArgTrpSerLysArgValTyrIleAlaTrpSerPheTrpIle 299
Db 783 CTAATGCTGACAGGCCGCTTTTCTACCGGATCTACGTAGCCTACTGTACTGTCTTACTGTC 842
Qy 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
Db 843 CTGGGACCACTTCTTCTATGCAAGATTCTCTTTGTTGGTTTCCAGCCGCTCTCTCATCA 902
Qy 320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPheLeuGlnTrp 339
Db 903 GAACACATGCCAGCTTTGGAGTGTGTGCTCTCTCATCATCTGCTTTCTGATGATTAC 962
Qy 340 LeuArg-AspArgLeuThrLysGlnGluPheGlnThrLeuPheLeuGlyValSerIle 359

```

```

Db 963 CTGGCCACACAAAGTTGAATCCACACCAATTCGAA---GTTTTTTCGGAGTGTATCTC 1019
Qy 359 uhaalaGlyAlaValPheLeuSerValIleTyrLeu---ThrTyrThrGlyTyrIleAl 378
Db 1020 CTTGGTGGCTTTCTCCCCCACTGGGAGCCTCTCCCATGCTAACAGGAAAAATTC 1079
Qy 378 aproTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIlePr 398
Db 1080 TCCCTGGACAGGCCGCTTTTACTCTCTGCTGAATCCCTCTAATGCTAATAAACATTC 1139
Qy 398 ollerleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspIle 418
Db 1140 CATTATTGCTGTTTCTGAGCACCAGCCACCACTGCTCTCTTCTTCTTCTTCTTCT 1199
Qy 418 uhisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAs 438
Db 1200 ACAGTCTCTTGTCTTCATGTTTCCAGTTGCCCTCTATTACTGCTTTTACCACCTGCTGA 1259
Qy 438 pgluargValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetVa 458
Db 1260 TGCTCGGATTTTATCATCATGATGATGGTGACAGCATGTACTTTTCAGCTGAATGGT 1319
Qy 458 largLeuMetLeuThrLeuThrProValValCysMetLeuSerAla---IleAlaPheSe 477
Db 1320 CGGTCTAATGCTGTGATGGCACCTGTTATGTGCTTCTTCTGGCATTTGGTGTTC 1379
Qy 477 rAsnValPheGluHisTyrLeuGlyAspMetLysArgGluAsnProProValGluAs 497
Db 1380 CAGGTGGCTGCTCCAC----- 1395
Qy 497 pSerSerAspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysVa 517
Db 1396 -----ATATATGAATAATCTCGACATAA 1418
Qy 517 largIysHisAlaThrGluGlnGlnLysThrGluGluGlyLeuGlyPro-----As 534
Db 1419 GTCCGCCACACAAAGAGCAGCAAGAGC-CAACAGGATCTTACTTACCTTATTAAGATGA 1477
Qy 534 nileYsserIleValThrMetLeuMetLeuMetLeuMetMetMetPheAlaValHisCy 554
Db 1478 GGTGGCGAGTGGGATGATGCTGGTCTTCTTCTCATCCTACCTACACGCTTTCATTC 1537
Qy 554 sThrTrpValThrSerAsnAla-TyrSerSerProSerValValLeuAlaSerTyrAsnH 574
Db 1538 GACTTGGGTGACAGTGAAGCCCTTATCTTCTCCCTCATTTCTGCTGCTGCTGGTGGT 1597
Qy 574 isAspGlyThrArgAsnIleLeuAspPheArgGluAlaTyrPheTrpLeuArgGlnA 594
Db 1598 GGGATGGCAGTAGGATCATTTTTCATGACITTCGAGAGCGGTATATTGCTCCGTCACA 1657
Qy 594 snThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyM 614
Db 1658 ATACTCCAGAGATGCAAAAGTCATGCTGTTGGGATATGGCTACCAATTTACTGCAAA 1717
Qy 614 etAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuV 634
Db 1718 TGGCAATTCGGACAATTTTAGTGACAATAACACATGCAATAATACCAATATTTCTCGAG 1777
Qy 634 aGlyLysAlaMetSerSerAsnGlu-ThrAlaAlaTyrLysIleMetArgThrLeuAsp 653
Db 1778 TAGGGCAGCAATGGCATCCACAGAAAGAACCTTATGAATCATGAGGGAGCTTGTAT 1837
Qy 654 ValAspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsn 673
Db 1838 GTGAGTATGTCTGTCTCATTTTGGAGGCCCTTACTGGGTATTTCTCGGATGATATCAAC 1897
Qy 674 LysPheLeuTrpMetValArgIle---AlaGluGlyGluHisProLysAspIleArgGlu 692
Db 1898 AAGTTTCTTGGATGGTCCGGATTGGAGGAGCAGACAGACAGACACATTAAGGAG 1957
Qy 693 SerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeu 712
Db 1958 AATGACTACTATCTCTACTCTGCTGGGGAATTCGTTGATCGTGGGCTTCTCCGCTGCTG 2017

```

```

QY 713 LeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPhe 732
Db 2018 CCACTGCTTATGACAAATGTTTACTACCGCTTTGGCAGGCTTACACAGAGCC 2077
QY 733 ArgThrProGlyPheAspArgThrArgAsnAlaGluLeuGlyAsnLysAspLys 752
Db 2078 AAGCGTCCACAGGCTTTGACCGCTTTCGAATGCTGAGATTGGTAATGAAGACTTTGAG 2137
QY 753 PheLysHisLeuGluGluAlaPheThrSerGluHisTyrPheLeuValAlaGlyTyrLysVal 772
Db 2138 CTTGATGCTCTGGAGAGCGTATACACAGACACTGCTAGTCAGGATATACAGGTA 2197
QY 773 LysAlaProAspAsnArg 778
Db 2198 AAGGACCTGGATAATCGA 2215

RESULT 11
LOCUS AY418286 2103 bp DNA linear GSS 17-DEC-2003
DEFINITION Mus musculus ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY418286
VERSION AY418286.1 GI:39774246
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 2103)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2103)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Source 1..2103
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>2103
/gene="ITM1"
/locus_tag="HCM6506"

ORIGIN
Alignment Scores:
Pred. No.: 1..15e-198 Length: 2103
Score: 2103.50 Matches: 405
Percent Similarity: 74.00% Conservative: 96
Best Local Similarity: 59.82% Mismatches: 142
Query Match: 47.80% Indels: 35
DB: 9 Gaps: 7

US-10-028-384-2 (1-826) x AY418286 (1-2103)

QY 105 TrpPheAsnTyrArgSerThrHisLeuAlaSerHisGlyPheTyrGluPheLeuAsn 124
Db 151 TACTTTAATATCGGACTACCGGTTCTCGCTCAGGAGGGGTTTATAAATCCATAAC 210
QY 125 TrpPheAspGluAlaTyrTyrProLeuGlyArgIleValGlyGlyThrValTyrPro 144

```

```

Db 211 TGGTTTGTATGACCGGGCTTGGTACCCCTTGGCGCAATCATTTGGAGGAAACAATTTACCCA 270
QY 145 GlyLeuMetIleThrAlaGlyLeuIleHisTyrPheLeuAsnThrLeuAsnIleThrVal 164
Db 271 GGTTTAATGATCACTCTCTGCTGCAATCTAGCATGTACTCCATTTCTTCCATATACATATT 330
QY 165 HistLeuArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSer 184
Db 331 GACATTCGGATGCTGTGTCTTCTGCGCCCACTTTCTCTCTTCCACACATCGTT 390
QY 185 ThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaCysPhe 204
Db 391 ACGTACCACCTTACCAAAGAGCTCAGGATGCGAGAGCTGGGCTTCTTGTGCTGCATG 450
QY 205 IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGly 224
Db 451 ATTGCTGTAGTTCTCTGGGTATATTTCTCGATCTGTAGCTGGCTCTTATGATATAGAGA 510
QY 225 IleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLysThrGly 244
Db 511 ATTGCTATCTTTTGCATGCTGCTTACTTACTACATGTGGATCAAGCAGTGAAGACTGGT 570
QY 245 SerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGly 264
Db 571 TCCATCTATTGGGCTGCCAAGTGTGCCCTGCTTATTTCTACATGCTCTTCTCATGGGA 630
QY 265 GlyTyrValPheIleLeuLeuIlePheLeuHisValPheValLeuLeuLeuMetGln 284
Db 631 GGCTATGTGTTCTCTGATCAACTTGAATCTCTCATGTCCTGTGTGTAATGCTGCACAGGC 690
QY 285 ArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuLeu 304
Db 691 CGTTTCTCACCAGACTACGTAGCTACTGTACTGTTTACTGCCCTGGCACCATTCTT 750
QY 305 SerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAla 324
Db 751 TCTATGCAGATTTCTGTTGTTTCCAGCCGCTCTCTCATCAGAACACATGGCAGCC 810
QY 325 AlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeu 344
Db 811 TTTGGAGTGTGTTGCTCTCTGTCAGATCCATGCTTTCGTAGATTACCTGGCAGCAGTTG 870
QY 345 ThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaVal 364
Db 871 AATCCACAGCAATTCGAAGTCTTTCCGGAGTGTATCTCCCTG---GTTGGCTTTGTC 927
QY 365 PheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPhe 384
Db 928 CTCCTCACTGTGGAGCTCTCCGCA-AAA-----ATTCTCCCTGGACAGGCGCTTTC 980
QY 385 TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleLeuAlaSerValSer 404
Db 981 TACTCTCTGCTGGATCCCTCTTATGCTAAGATAACATATCCCATTTATGTCATCTTCT 1040
QY 405 GluHisGlnProThrThrTrpValSerPhePheAspLeuHisIleLeuValCysThr 424
Db 1041 GAGCACAGCCCAACACTGCTTCTTACTATTTGATCTACAGCTCTCTGCTTCATG 1100
QY 425 PheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAla 444
Db 1101 TTTCCAGTTGGCTCTTATTACTGCTTTAGCAACCTGCTGCTGATCTCGGATTTTATCATC 1160
QY 445 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 464
Db 1161 ATGTATGTTGGACGACGATGTTCTTTTCAGCTGTATGTTGGTGTCTTATGCTGGTATTG 1220
QY 465 ThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeu 484
Db 1221 GCACCTGTTATGTGCAATCTCTTCTGGCATTTGGTGTTCCTCCAGGTGTGTGCCATATG 1280
QY 485 GlyAsp---AspMetLysArgGluAsnProProValGluAspSerSerAspGluAspAsp 503

```

Db 1291 AAAAATCTGGACATACAGTCCGCCA----- 1304
 Qy 504 LysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyValArgLysHisAlaThrGlu 523
 Db 1305 -----GACAGAAGAGCAAGAACAGGATCTACTTAC 1340
 Qy 524 GlnGluValThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMet 543
 Db 1341 CCTATTAGAAATGAG-----GTGGCAGTGGGATGATCTACTGTGATG 1382
 Qy 544 LeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSer 563
 Db 1383 GCTTTTCTCATCACTACACGCTTATTCGACITGGTGCACAGTGAAGCCTATTCT 1442
 Qy 564 SerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAsp 583
 Db 1443 TCTCCCTCCATGTGACTGTCTGCTGGTGGGATGGCAGTAGGATCAATTTTGTATGAC 1502
 Qy 584 PheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSer 603
 Db 1503 TTCGAGAAGCGTATTATTGGCTCCGTCACATATCCAGAGGATGCCAAAAGTCATGTCA 1562
 Qy 604 TrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsn 623
 Db 1563 TGGTGGATATGCTACCAATATTCTGCAATGGCAATCGGACATTTTATGTGACAA 1622
 Qy 624 AsnThrTrpAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThr 643
 Db 1623 AACACATGGAATATATACCATATTCTCGAGTAGGCGAGCAATGGCATCCACAGAAGA 1682
 Qy 644 AlaAlaTyrLysIleMetArgThrLeuAspValAspTyrPheThrProGlnGlyGluPhe 702
 Db 1683 AAACCTATGAATCATGAGGAGCTTGATGTGAGCTATGTCTTGTCAATTTTGGAGGC 1742
 Qy 664 ValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIle---Ala 682
 Db 1743 CTACTTGGTATTCTTCGGATGATATCAACAAGTTTCTTTGGATGTCGGATTGGAGA 1802
 Qy 683 GluGlyGluHisProLysAspIleArgGlySerAspTyrPheThrProGlnGlyGluPhe 702
 Db 1803 AGCAGACAGAGAGAGACACATTAGAGAGATGACTACTACTCTCTACTCGGGAATC 1862
 Qy 703 ArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyr 722
 Db 1863 CGTGTGATCTGAGGGTTCTCGGTCGCTGCTCAACTGCTTATGTACAAAATGTGTAC 1922
 Qy 723 TyrArgPheGlyGluMetGlnLeu-AspPheArgThrProProGlyPheAspArgThrAr 742
 Db 1923 TACCGCTTTGG-----CAGGCTACACAGAAGCGTCCACAGGTTTGACCGGTTCG 1976
 Qy 742 gAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSe 762
 Db 1977 AAATGCTGAGATTGGTAATAAAGACTTGGCTTGGATGCTCTCGAGAAAGCGGTATACAC 2036
 Qy 762 rGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArg 778
 Db 2037 AGAACCTGGCTAGTACAGTATACAGGTAAGGACCTGGATATCGA 2085

CNS0ADYZ 2419 bp mRNA linear HTC 06-FEB-2004
 Arabidopsis thaliana Full-length cDNA Complete sequence from clone
 GSUTPGH512C09 of Hormone Treated Callus of strain col-0 of
 Arabidopsis thaliana (thale cress).
 BX316490
 BX316490.1 GI:42474670
 HTC; GSUT cDNA.
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 2419)

AUTHORS Castelli, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C., Menard, M., Cruaud, C., Quetier, F., Scarpelli, C., Schachter, V., Temple, G., Caboche, M., Weissenbach, J., and Salanoubat, M.
 TITLE Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2419)
 AUTHORS Direct Submission
 TITLE Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 JOURNAL The sequences are based on single pass reads.
 COMMENT Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.
 Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M.
 URGV INRA : Clepet C., Caboche M.
 Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences). 5 prime and 3 prime are assembled with Phrap.
 http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full-length
 http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
 FEATURES
 source
 1. 2419
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Col-0"
 /db_xref="taxon:3702"
 /clone="GSUTPGH512C09"
 /tissue_type="Hormone Treated Callus"
 /plasmid="pCMVSPORT_6"
 1. 2419
 /gene="At1g34130"
 ORIGIN
 Alignment Scores:
 Pred. No.: 9,11e-195 Length: 2419
 Score: 2065.50 Matches: 404
 Percent Similarity: 68.40% Conservative: 109
 Best Local Similarity: 53.87% Mismatches: 188
 Query Match: 46.93% Indels: 49
 DB: 3 Gaps: 12
 US-10-028-384-2 (1-826) x CNS0ADYZ (1-2419)
 Qy 60 GlyGlyLeuSerGlnProAlaGlyTrpGlnSer----- 70
 Db 114 GGAGGAAATATCAGACCCAGCAATCCGAATCCATGCGCAGAACCCAGATCTACTCAAC 173
 Qy 71 -----LeuLeu 72
 Db 174 ACTAGTTTCTTCTCATTCATTAATCCCTAAATTTGAAGCAGAAACACACAGAGCTTCTACTC 233
 Qy 73 SerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerArgLeuPheAlaVal 92
 Db 234 CGTATCTCGATCTCTGGTCTCGTCTACATCTTAGCTTTTCATAGCTCTCTCTCAGTGT 293
 Qy 93 IleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgSerThrHis 112
 Db 294 CTCGGTACGATCAATGATCCAGCAATTCGATCCGATTTTCACTACCTACCTACGACTCT 353
 Qy 113 HisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyr 132
 Db 354 TTCCTCAGAGAAAGGCTTCTACAGTTTGGAACTGGTTCGATTCGAGAGTGTGTAT 413
 Qy 133 ProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrAlaGlyLeu 152
 Db 414 CTTCTAGTTCGAATCATCGGGGGTACTCTTTATCCGGTCTTATGTGTGTCAGCTCGCTTA 473

gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2. (bases 1 to 2079)
 AUTHORS Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
 Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cartgill, M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 source
 1. 2079
 Location/Qualifiers
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 <1..2079
 /gene="TIM1"
 /locus_tag="HMC650s"
 ORIGIN
 Alignment Scores: Length: 2079
 Pred. No.: 2,04e-157
 Score: 1691.00
 Matches: 333
 Percent Similarity: 61.19%
 Best Local Similarity: 49.70%
 Mismatches: 228
 Query Match: 38.42%
 Indels: 32
 DB: 9
 Gaps: 3
 US-10-028-384-2 (1-826) x AY418285 (1-2079)
 QY 106 PheAsnTyrArgSerThrHisLeuAlaSerHisGlyPheTyrDluPheLeuAsnTrp 125
 DB TTTAAATATCGGACTACCAAGNNCCCTGGCTGNNNNNGGGTTTATAAATTCATACATGG 213
 QY 126 PheAspGluArgAlaTyrProLeuGlyArgGlyValGlyGlyThrValTyrProGly 145
 DB TTTGATGACCGAGCGTGTACCTTTGGGACGAATCATTTGGAGGAGCANNNTTACCAGGT 273
 QY 146 LeuMetIleThrAlaGlyLeuIleHisTyrPheLeuAsnThrLeuAsnIleThrValHis 165
 DB TTAATGATACCTCTGCTGCAATCTACCATGACTCCATTTTCCACATCACCATCGAC 333
 QY 166 IleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThr 185
 DB ATTCCGAATGCTNTGTGTCTCTGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 393
 QY 186 PheLeuLeuThrArgGluLeuTyrAsnGinglyAlaGlyLeuLeuAlaAlaCysPheIle 205
 DB TACCACCTTACCAAGAGCTCAAGATGAGGAGGCTGGGCTTCTTCTGCTGCTGCTGCT 453
 QY 206 AlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIle 225
 DB GCTGTAGTCTCTGGATATATCTCCGATCTGTGGCTGGCTCTATATATATATATATAT 513
 QY 226 AlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValLysSerValIleThrClySer 245
 DB GCCATCTTTTGGCACTACTACCTACTACTGATGATGATGATGATGATGATGATGATG 573
 QY 246 ValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGly 265
 DB ATCTGTGGGAGCAGTAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 633
 QY 266 TyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArg 285
 DB TATGTGTCTGATCACTTAATCTCTCCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTG 693
 QY 286 TyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSer 305
 DB TTCTCTCACCAGGATCTATGTGGCTACTGTACTGTCTTACTGCTGCTGCTGCTGCTTCT 753

QY 306 MetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaA 325
 DB ATGCAGATCTCTTTGTGGTTTCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 813
 QY 326 GlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThr 345
 DB NNN 873
 QY 346 LysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPhe 365
 DB CCANPACAAATTNNNGTTCTTTTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 933
 QY 366 LeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProThrPsrGlyArgPheTyr 385
 DB NNN 933
 QY 386 SerLeuTrpAspThrGlyTyrAlaIleHisIleProIleIleAlaSerValSerGlu 405
 DB TACTGCTGGATCTCTTATGCTTAAGAACCAATCCCATCATCTTCTGTGCTGAG 1053
 QY 406 HisGlnProThrThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPhe 425
 DB CATCAGCCCAACCTGCTCTCATCTATTTTGCACCTNNAGCTCTCTGCTCTCATGTT 1113
 QY 426 ProAlaGlyLeuTrpPheCysIleLeuAsnIleAsnAspGluArgValPheValAlaLeu 445
 DB CAGTTGGCTCTATTAATGCTTTNGCACTGCTGATGCCCGGATTTTATCATCATG 1173
 QY 446 TyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThr 465
 DB TATGCTGTGACCAAGCATGCTACTTTTCACTGCTTAATGNNNNNNNNNNNNNNNN 1233
 QY 466 ProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGly 485
 DB NNN 1290
 QY 486 AspAspMetLysArgGluAsnProValGluAspSerSerAspGluAspAspLysArg 505
 DB NNN 1290
 QY 506 AsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGlu 525
 DB NNN 1291
 QY 526 LysThrGluGluGlyLeuGlyPro-----AsnIleLysSerIleValThrMetLeu 542
 DB AAGNNNNAGGATTCGCCNCCCTATTAAAGAAATGAAGTGGCAAGTGGGATGATGCTG 1389
 QY 543 MetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyr 562
 DB ATGGCTTTCTTCTCATCTACCTACCTTTCATTCACCTGGGTGGTACGAGGCTAC 1449
 QY 563 SerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAsp 582
 DB TCTTCTCGCTCATCTGATCTATCTGCGGCTGGTGGGATGCGAGTAGGATCATATTGAT 1509
 QY 583 AspPheArgGluAlaTyrPheThrLeuArgGlnAsnThrAspGluHisAlaAspValMet 602
 DB GACTTCCGAGAGCATATATGCTGCTCACAATCTCCAGAGGATGCGAGGCTCATN 1569
 QY 603 SerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAsp 622
 DB NCTGCTGGGATTTATGGCTATCAGATATGCAATGCAACCCGACAAATTTAGTGGAC 1629
 QY 623 AsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGlu 642
 DB AATAACATGGAATATACCATATTTCTNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1689
 QY 643 ThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGly 662
 DB GAAAAGGCTATGATCATGAGGAGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCT 1749
 QY 663 GlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrpMetValArgIle--- 681

Db 1750 GGCCTCAGCGGTATCTCTCATGATATCAACAGTTCNNNGATGGTCCGATGGA 1809
 QY 682 AlaGluGlyGluHisProLysAspLeuArgGluSerAspTyrPheThrProGlnGlyGlu 701
 Db 1810 GGGAGNACAGANNANGCAACATATCAAGGAGATGACTATATATATCCACTGGGGAG 1869
 QY 702 PheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSer 721
 Db 1870 TTCGGTGTGNNCNTGAAGGTTCTCCAGTGTCTCCANNNNNNNNNNNNNNNNNN 1929
 QY 722 TyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArgThr 741
 Db 1930 NNN 1989
 QY 742 ArgAsnAlaGluLeuGlyAsnLysAspLeuLysHisLeuGluGluAlaPheThr 761
 Db 1990 CGAAATGCTGAGATGGGAATAAGACTTTGAGCTTGATGCTCGGAGGAGCATATACC 2049
 QY 762 SerGluHisTyrLeuValArgIleTyrLys 771
 Db 2050 ACAGAACATGCTGCTGAGGATATACAG 2079

RESULT 14
 BUS15954
 LOCUS
 DEFINITION AGENCOURT 10120433 NIH MGC 134 Mus musculus cDNA clone
 IMAGE:6512642 5', mRNA sequence.
 BUS15954
 VERSION BUS15954.1 GI:22823480
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 950)
 NIH-MGC <http://mgi.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-rc@mail.nih.gov
 Tissue Procurement: Dr. David Rowe
 cDNA Library Preparation: Invitrogen Corp
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Cloned by: Agencourt Bioscience Corporation
 Cloned distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM14085 row: b column: 03
 High quality sequence stop: 680.
 Location/Qualifiers
 1..950
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6512642"
 /tissue_type="undifferentiated limb"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 134"
 /note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
 Cloned unidirectionally. Primer: Oligo dT. Average insert
 size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
 this is a NIH_MGC Library."

FEATURES
 source
 Alignment Scores:
 Pred. No.: 3,71e-145 Length: 950
 Score: 1563.00 Matches: 300
 Percent Similarity: 96.23% Conservative: 6
 Best Local Similarity: 94.34% Mismatches: 10
 Query Match: 25.51% Indels: 2
 DB: 5 Gaps: 0

ORIGIN

RESULT 15
 CR610915
 LOCUS
 DEFINITION full-length cDNA clone CS0DK0101001 of HeJa cells Cot 25-normalized
 of Homo sapiens (human).
 CR610915
 ACCESSION
 VERSION CR610915.1 GI:50491722

US-10-028-384-2 (1-826) x BUS15954 (1-950)
 QY 215 SerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 234
 Db 1 TCAGTGGCGGATCTTTGATATGAAGGATGGCCATTTTGGCTTCAGTTCACTTAC 60
 QY 235 TyrLeuTyrValLysSerValLysThrGlySerValPheThrMetCysCysLeu 254
 Db 61 TACTTATGGTAAAGTCTGTGAAGACCGGCTGTGTCTTGGCAATGTGCTGCTGCTG 120
 QY 255 SerTyrPheTyrMetValSerAlaTyrGlyTyrValPheIleIleLeuLeuIlePro 274
 Db 121 TCATATTTTACATGGTCTCTGCTGGGAGGTATGTGTTCATCATCACTCATCCCT 180
 QY 275 LeuHisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyr 294
 Db 181 CTCATGT 240
 QY 295 SerThrPheTyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPheGln 314
 Db 241 AGCATTTCATCATTTGGGTAAATATTCATGCAGATACCTTTTGTGGATTTGAG 300
 QY 315 ProIleArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyr 334
 Db 301 CCAATCAGAACAGCGAGCACATGGCAGCTGCAGGTGTCTTTCGCTGTGCAAGCTTAC 360
 QY 335 Ala-PheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePh 354
 Db 361 GCTTTTTCAGATATCTGAGACCGGTGACAAACAGGAGTTCAGACCCCTTTCTT 420
 QY 354 eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrIleThrTyrTh 374
 Db 421 TTTGGGTGTCTCACTAGCTGAGCGGTGTGTTCCTTAGTGTCTATCTATCTGACATACAC 480
 QY 374 rGlyTyrIleAlaProTyrSerGlyArgPheTyrSerLeuTyrAspThrGlyTyrAlaL 394
 Db 481 AGTTATATTCACCATGGAGTGGCAGGTTTATTTACTATGGGATCTGGGTATGCAAA 540
 QY 394 sIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrPheValSerPh 414
 Db 541 AATACATTCACATTTATTCATCAGTGTCTGAACATCAGCTTACGACATGGGTGTCTTT 600
 QY 414 ePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTyrPheCysIleL 434
 Db 601 CTTCTTGTATCATATATCTTGTATGTACTTCCAGCAGGCTATGTTCTGCAATCAA 660
 QY 434 sAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAl 454
 Db 661 AAATATCAACGATGAAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
 QY 454 aGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaI 474
 Db 721 CNGAGTGTATGGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
 QY 474 eAlaPheSerAsnValPheGluHisTyrLeuGlyAspAspMetLysArgGluAsnProPr 494
 Db 781 CGCTCTCTCCAAATGTTTTTGGACACTATTTGGGGGATGACATTTGAAGGAAACCCACC 840
 QY 494 oValGluAspSerSerAspGluAspLysArgAsnGlnGlyAsnLeuTyrAspLysAl 514
 Db 841 TGTGGAGGACGACACTGATGATGATGACAAAGAAACCCCGGAAACTTTGTATGACAAAGCC 900
 QY 514 aGlyValValArgLysHisAlaThrGluGln-GluLysThrGluGluGly 530
 Db 901 AGGTAAAGTGAAGAAGCATGTGCCAGACAAAGAGAAACCTGCAACGAGGA 950

CR610915 2212 bp mRNA linear HTC 21-JUL-2004
 full-length cDNA clone CS0DK0101001 of HeJa cells Cot 25-normalized
 of Homo sapiens (human).
 CR610915
 ACCESSION
 VERSION CR610915.1 GI:50491722

KEYWORDS HTC; CNSLT CDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2212)
 AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Paraday Avenue
 REFERENCE 2 (bases 1 to 2212)
 Genoscope.
 AUTHORS Direct Submission
 TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 FEATURES
 source location/Qualifiers
 1..2212
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DK010Y001"
 /tissue_type="Hela cells Cot 25-normalized"
 /plasmid="pCMVSPORT_6"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,43e-139 Length: 2212
 Score: 1513.00 Matches: 282
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 34.38% Indels: 0
 DB: 3 Gaps: 0
 US-10-028-384-2 (1-826) x CR610915 (1-2212)
 QY 545 MetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSer 564
 DB 3 ATGCTATTGATGATGTTTGTCTCCACTGTACTGGTGCACAGCAATGCCCTACTCTAGT 62
 QY 565 ProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspPhe 584
 DB 63 CCAAGTGTAGTCTCGGCTCATACATCATGATGGCCAGGAAATATCTTAGATGATTTT 122
 QY 585 ArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrp 604
 DB 123 AGAGAAGCTTACTTTTGGCTAAGCAAAATACAGATGACATGCCAGTAAATGTTCTGG 182
 QY 605 TrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsn 624
 DB 183 TGGGATTATGGCTATCAGATAGCTGGAAATGGCTAATAGAACTACGTTGGTGGATATAAC 242
 QY 625 ThrTrpAsnSerHisIleAlaLeuValGlyValAlaMetSerSerAsnGluThrAla 644
 DB 243 ACCTGGAAATACACCCACATAGCACTGGTGGGAAAGCTATGCTCTTAATGAAACAGCA 302
 QY 645 AlaTyrIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyVal 664
 DB 303 GCTATAAAATCATCAGGACTCTAGATGTAGATTATGTTTGGTATTTTTCGAGGGTT 362
 QY 665 IleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGly 684
 DB 363 ATTGGCTATTCTGGTGTATGATATCAACAATTTCTCTGGATGGTAGTAGCTGAAGGA 422
 QY 685 GluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArgVal 704

Db 423 GAACATCCCAAGACATTCGGGAAAGTGACTATTATTTACCCCAAGGAGAAATCCGTGTA 482
 QY 705 AspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArg 724
 Db 483 GACAAAGCAGGATCCCTACTCTTTGTTGATTTGCCTTATGTATATAAATGTCATACTACAGA 542
 QY 725 PheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsnAla 744
 Db 543 TTTGGAGAAATCGAGCTGGATTTTCGTACACCCCCAGGTTTTCACCGAACACACGTAATGCT 602
 QY 745 GluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHis 764
 Db 603 GAGATTGGNAATAAGGACATTAAATTCAACATTTGGAGAGGCTTTTATCATCAGAACAC 662
 QY 765 TrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLys 784
 Db 663 TGGCTTTGTAGGATATATAAAGTAAAGCACCTGATAACAGGGAGACATTAGATCACAA 722
 QY 785 ProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArg 804
 Db 723 CTTGAGGTCCACCAACATTTTCCCAAAACAGAGTATTTGTCAAGAGAGACTACCAAAGG 782
 QY 805 LysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLys 824
 Db 783 AAGCGTGGCTACATTAAAAATAAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAGAAG 842
 QY 825 ThrVal 826
 Db 843 ACTGTT 848

Search completed: December 15, 2004, 11:49:55
 Job time : 6518.33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 / Search time 7954.53 Seconds
(without alignments)
4268.513 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 718

Sequence: 1 MGSDSCVLSVFQILKVI.....TRSTKTRRSIKRPELGLRV 718

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues.

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131645_6800/app_query.fasta_1.3740
-DB=GenEmbl -QFMT=fastap -SURFIX=Oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTPMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10028384 @CEN 1 1 20262 @runat_14122004_131645_6800 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	718	100.0	2733	6	AX799086 Sequence
2	718	100.0	2733	8	D28952 Saccharomyc
3	631	87.9	3633	8	Y22544 S.cerevisia
4	605	84.3	2000	6	AX595714 Sequence

5	605	84.3	2000	6	AX819898 Sequence
6	605	84.3	2000	6	AX830928 Sequence
7	121	16.9	381	6	C0450623 Sequence
8	71	9.9	3447	8	SCALK1 X87672 S.cerevisia
9	39	5.4	110000	8	CR382123_09 Continuation (10 o
10	39	5.4	110000	8	CR382123_10 Continuation (11 o
11	35	4.9	300029	8	AE016903 Ereothec
12	32	4.5	2256	6	AX489023 Sequence
13	32	4.5	5048	8	AB015232 Schizosac
14	32	4.5	40907	8	AB015232 S.pombe c
15	32	4.5	110000	8	SPBC1271 Continuation (4 of
16	32	4.3	110000	8	CR382134_03 CR380947_0
17	29	4.0	110000	8	CR382128_21 Continuation (22 o
18	20	2.8	2417	3	AF132552 Drosophil
19	20	2.8	2417	6	AX799088 Sequence
20	20	2.8	2699	6	CQ589353 Sequence
21	20	2.8	2953	6	CQ596819 Sequence
22	20	2.8	4922	6	CQ589352 Sequence
23	20	2.8	75650	2	AC018145 Drosophil
24	20	2.8	162921	3	AC007853 Drosophil
25	20	2.8	181132	3	AC008206 Drosophil
26	20	2.8	227219	3	AE003750 Drosophil
27	19	2.6	94169	8	EX842635 Neurospor
28	17	2.4	297	8	AF548377 Talaromyc
29	17	2.4	558	6	AX079689 Sequence
30	17	2.4	1664	6	CQ782554 Sequence
31	17	2.4	1664	6	BD127193 Primer fo
32	17	2.4	1664	9	AK074587 BC003206 Mus muscu
33	17	2.4	1734	10	BC003206 Mus muscu
34	17	2.4	2481	6	AX799082 Sequence
35	17	2.4	2481	9	AY074880 Homo sapi
36	17	2.4	2710	6	AX799084 Sequence
37	17	2.4	2855	6	CQ573675 Sequence
38	17	2.4	4236	10	BC052433 Mus muscu
39	17	2.4	6153	6	CQ573674 Sequence
40	17	2.4	18405	2	AC017914 Drosophil
41	17	2.4	54118	3	U13019 Caenorhabdi
42	17	2.4	166908	3	AC104631 Drosophil
43	17	2.4	170691	9	AC104643 Homo sapi
44	17	2.4	176428	5	EX255961 Zebrafish
45	17	2.4	177253	9	AC092024 Homo sapi

ALIGNMENTS

RESULT 1	AX799086	Sequence 5 from Patent WO03054008.	2733 bp	DNA	linear	PAT 08-OCT-2003
LOCUS	AX799086	Sequence 5 from Patent WO03054008.	2733 bp	DNA	linear	PAT 08-OCT-2003
DEFINITION	AX799086	Sequence 5 from Patent WO03054008.	2733 bp	DNA	linear	PAT 08-OCT-2003
ACCESSION	AX799086	Sequence 5 from Patent WO03054008.	2733 bp	DNA	linear	PAT 08-OCT-2003
VERSION	AX799086.1	GI:37605060				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
1						
Perreault, C. and McBride, K.						
Mammalian SIMP protein, gene sequence and uses thereof in cancer						
therapy						
Patent: WO 03054008-A 5 03-JUL-2003;						
Comptatigene Inc. (CA)						
Location/Qualifiers						
1..2733						
/organism="Saccharomyces cerevisiae"						
/mol_type="genomic DNA"						
/db_xref="taxon:4932"						
/note="STT3, Acc# D28952"						

ORIGIN

Alignment Scores:	0	Length:	2733
Pred. No.:	718.00	Matches:	718
Score:			

Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-10-028-384-6 (1-718) x AX799086 (1-2733)			
Qy	1	MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle	20
Db	219	ATGGGATCCGACCGGCTGGTGGTTTGTCTGTGTGTTTCAGACCACTCCTCAAGCTCGTCATC	278
Qy	21	PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu	40
Db	279	TTCTGGGCGAATTTTGGGCTGCCAATCATCATCAGTTTGTTCAGCATCAAAATTTGAG	338
Qy	41	SerIleIleHisGluPheAspProTTPheAsnTyrArgAlaThrLysTyrLeuValAsn	60
Db	339	TCTATTATCCATGAATTCGACCCCTGGTTCATATATAGGCTACCAAAATATCTCGTCAAC	398
Qy	61	AsnSerPheTyrLysPheLeuAsnTTPheAspAspArgThrTTPyrProLeuGlyArg	80
Db	399	AATTCGTTTACAAAGTTTGAACCTGTTTGACGCGCTACCTGGTACCCCTCCGAAGG	458
Qy	81	ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla	100
Db	459	GTTACTGGAGGACCTTTATATATCTCTGTTTGTATGACGACTAGTGGTTCATCTGGCACGCC	518
Qy	101	LeuArgAsnTTPheLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro	120
Db	519	CTGGCAACTGGTTGGGCTTGCCCATTTGACATCAGAAACGTTTGTGTGTATTTGGGCA	578
Qy	121	LeuPheSerGlyValThrAlaTTPAlaThrTyrGluPheThrLysGluIleLysAspAla	140
Db	579	CTATTTCTCGGGTCACCGCTGGCGACTTACGAAATTTACGAAAGAGATTAAGATGCC	638
Qy	141	SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer	160
Db	639	AGCGCTGGGCTTTTGGCTGCTGGTTTATAGCAATTTGTCCTGGTTATATCTAGATCA	698
Qy	161	ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet	180
Db	699	GTGGCGGGCTCTACGATAATGAGCCATTTGCCATTTACACTATTAATGGTCTCTTTTCATG	758
Qy	181	PheTTPileLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe	200
Db	759	TTTGGATTAAGGCCCAAGACTGGCTCTATCATGCACGCAACGTTGCAGCTTTATTC	818
Qy	201	TyrPheTyrMetValSerAlaTTPGlyGlyTyrValPheIleThrAsnLeuIleProLeu	220
Db	819	TACTTCTACATGGTGGCTGGGCTGGGATGATGGTGTTCATACCAACTTGATCCCACTC	878
Qy	221	HisValPheLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr	240
Db	879	CATGCTTTTGTGATTTTGAUGGGCAGATATTCGTCGCAACTGTTATCTGCCTACACC	938
Qy	241	ThrTTPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro	260
Db	939	ACTTGGTACGCTATTTGAACCTGTTGCATCCATCAGATCCCATTTTCGCTTCTCTACCT	998
Qy	261	IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla	280
Db	999	ATCAGGTCTACAGCACCATGGCCGCTATGGGTGGTTTTCGTTTGTATTCAGATTTGCGCC	1058
Qy	281	PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal	300
Db	1059	TTCCGTGACTTCTGGTGAAGGGCCAAATCAGCACAGCTTAAGTTTAAAGCTCATCATGATGTT	1118
Qy	301	SerLeuPheLeuLeuLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly	320
Db	1119	TCTCTGTTTGTGATCTTGGTCTTGGTGGTGGTGGGACTTCTCTGCTTACCTATATGGG	1178
Qy	321	LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTTPAspThrAsnTyrAlaLysIle	340
Db			
Db	1179	TTGATTGCCCTTGGACTGGTAGATTTTATTGGTTATGGGATACCAACTACGCAAGATC	1238
Qy	341	HisIleProIleIleAlaSerValSerGluHisGlnProValSerTTPProAlaPhePhe	360
Db	1239	CACATTTCTATCATTTGGCTCCGTTTCCGAACATCAACCCGTTTCTGGTCCGCTTTCTTC	1298
Qy	361	PheAspThrHisPheLeuIleTTPhePheProAlaGlyValPheLeuLeuPheLeuAsp	380
Db	1299	TTTGATACCCACTTTTGTATCTGGCTATTTCCGCCCGGTGTATTTCTTACTATTCTCTGC	1358
Qy	381	LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly	400
Db	1359	TTGAAAGACGACGACGTTTGTGTCATCCGTTACTCCGTTCTGTCTGTCTGTTGCGGT	1418
Qy	401	ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla	420
Db	1419	GTTATGGTTAGATTGATGTTGACCTTTGACACGATCATCTGTGTGTCGCCGCGCTGCA	1478
Qy	421	LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle	440
Db	1479	TTGTCCAGATATTTGACATCTACCTGGATTTCAAGACAAGTGACCCGCAATACGCCATC	1538
Qy	441	LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr	460
Db	1539	AAACCTGCGGCACTACTGGCCAAATTTGATTTTCCGGATCATTCATCTTTTATTTGAT	1598
Qy	461	LeuPheValPheHisSerThrTTPValThrArgThrAlaTyrSerSerProSerValVal	480
Db	1599	CTTTTCGCTTCCATTTCTACTTGGGTAAACAAGACTGCATACTCTTCTCTCTCTGTTT	1658
Qy	481	LeuProSerGlnThrProAspGlyLysLeuAlaIleLeuLeuAspPheArgGluAlaTyr	500
Db	1659	TTGCCATCAACAACCCAGATGGTAAATTTGGCGTTGATCGACGACTTCAGGGAAGCGTAC	1718
Qy	501	TyrTTPLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTTPAspTyrGly	520
Db	1719	TATTTGGTTAAGAAATGAACCTCTGATGAGGACAGTAAGGTTGCAGCGTGTGGATTACGGT	1778
Qy	521	TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTTPAsnAsn	540
Db	1779	TACCAAAATTTGGTGGATGGCAGACAGAACCACTTTAGTCGATAACAACACGTTGAAACAAT	1838
Qy	541	ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle	560
Db	1839	ACTCACATCGCCATCGTTGGTAAAGCCCATGGCTTCCCTCCCTGAAGAGAAATCTTACGAAAT	1898
Qy	561	LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly	580
Db	1899	CTAAAAGACATGATGTCGATTTATCTTGGTCATCTTTTGGTGTCTTAATTTGGGTTTGGT	1958
Qy	581	GlyAspAspIleAsnLysPheLeuTTPMetIleArgIleSerGluGlyIleTTPProGlu	600
Db	1959	GGTGTATGACATCAACAATTTCTTGTGGATGATCAGAAATAGCGGGAATCTTGCCGCAAA	2018
Qy	601	GluIleLysGluArgTyrPheTyrThrAlaGluGlyGlyTyrArgValAspAlaArgAla	620
Db	2019	GAGATAAAAGACGCTTATTTCTATACCGCAGGAGAGAAATACAGAGTAGATGCAAGGCT	2078
Qy	621	SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu	640
Db	2079	TCTGACACCATGAGGAATCTCGCTACTTTTACAAGATGCTCTACAAGATTTTCCCAACAATTA	2138
Qy	641	PheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal	660
Db	2139	TTCAATTTGGTGGCCAGCCACTGACAGAGTGGCTCAACAAATGATCACCATTAGACGTC	2198
Qy	661	ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTTPMetValArgIleTyr	680
Db	2199	CCACCATTTAGACTACTTCCGACGAGTTTACTTCCGAAAACTGGATGGTTAGAAATATAT	2258
Qy	681	GlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg	700
Db	2259	CAATTGAAGAGATGATGCCCAAGTAGAATCTTTGAGGGGACGTTGTGAGTTAACCCAGG	2318

Qy 701 SerSerThrIysThrArgSerIleLysArgProGluLeuGlyLeuArgVal 718
 Db 2319 TCTTCTACGAAACCAAGAGGTCCTAAAGAGACCTGAATAGGCTTGAAGATC 2372

RESULT 2
 YSCSTT3 2733 bp DNA linear PLN 18-JAN-2003
 LOCUS
 DEFINITION Saccharomyces cerevisiae gene for STT3 protein, complete cds.
 ACCESSION D28952
 VERSION D28952.1 GI:468493
 KEYWORDS
 SOURCE
 ORGANISM
 Saccharomyces cerevisiae (baker's yeast)
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE
 1 Yoshida, S., Ohya, Y., Nakano, A. and Anraku, Y.
 STT3, a novel essential gene related to the PKC1/STT1 protein
 Kinase pathway, is involved in protein glycosylation in yeast
 Gene 164 (1), 167-172 (1995)
 96060859
 7590309
 2 (bases 1 to 2733)
 YOSHIDA, S.
 Direct Submission
 Submitted (14-MAR-1994) Satoshi Yoshida, Faculty of Science,
 University of Tokyo, Department of Biology, 7-3-1, Hongo,
 Bunkyo-Ku, Tokyo 113, Japan (Tel:81-3-3812-2111 (ex.4461),
 Fax:81-3-3812-4929)

FEATURES
 source
 1. .2733
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /strain="S288C"
 /db_xref="taxon:4932"
 /clone_lib="pST3-105"
 /clone_lib="library of T. Yoshida and Y. Anraku"
 2.9. .2375
 /genes="STT3"
 219. .2375
 /genes="STT3"
 /codon_start=1
 /product="STT3 protein"
 /protein_id="BA08079.1"
 /db_xref="GI:603808"
 /translation="WGSDRSCVLSVFQTLKLVFVAFGAAISSLFVAVIKPESIIH
 EFDPNFYRATKLYNNGFYKFLNWFDDRTWYPLGRVTGTLPLGLMTGAFIWHALR
 NMLGPIFDIRNVCVLFAGLFGVTAWATFEYKEIKDASGLLAAGFIAIVPGYISRS
 VAGSYDNEAIAITLLMVTFMFWIKAKQKTSIMHATCAALFYFVWSWAGGYVFTNLI
 PLHVFLLILMGYSKLYSATYTTVAIGTVASMOI PFVGPLPISNDHMAALGVFGLI
 QIVAGDFVKGQISTAKSKVIMWSLFIULVLGVGLSALTVMGLIAPWTGRFYSGLWD
 TNYAKIHPIIASVSEHQVSWPAFFEDTHFLIWLFPAGVFLFLDLKDHVFIAYS
 VICSIFAGVNRMLTLPVICSAVALSKIFDIYDFKTSRDKYAKPAALIAKLI
 VSGSFIYLLFVFTSWTRTAYSSVPLPSQTPDGKUALIDDREAYVILWRMSD
 ESKVAWMDYQYQGMADRTLLVDNNWNTHIAIVKAMASPEKSEYILKEHDV
 DYVLVIFGLIGDDDKINKLMIRI SEGIPPEIKERYFYTAEGEYRVDARASETM
 RNSLYKNSKDFPOLFNQGQATDRVQOMITPLDVPFLDYFDEFTSENWVRIVQL
 KKDDAQCRILRDVGLSTRSSKTRRSIKRPLGLRV"

ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 2733
 Score: 718.00 Matches: 718
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x YSCSTT3 (1-2733)

Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValle 20
 |||

Db 219 ATGGATCCGACCGTCGTGTGTTTGTCTGTGTTTTCAGACCATCTCAAGCTCGTCATC 278
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 |||
 Db 279 TTCGTGGCGAATTTTGGGGCTGCCATATCATCATCGTTTGTGTCAGTCATCAAAATTGAG 338
 |||
 Qy 41 SerIleIleHisGluPheAspProTTPPheAsnTyrArgAlaThrLysLysTyrLeuValAsn 60
 |||
 Db 339 TCTATTATCCATGAATTCGACCCCTGGTTCATATTATAGGCTACCAAAATATCTCGTCAAC 398
 |||
 Qy 61 AsnSerPheTyrLysPheLeuAsnTTPPheAspAspArgThrTTPyrProLeuGlyArg 80
 |||
 Db 399 AATTCGTATTTACAAGTTTTTGAACCTGTTTGGACGACCGTACCTGGTACCCCTCGAAGG 458
 |||
 Qy 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla 100
 |||
 Db 459 GTTACTGGAGGGACTTTATATCTCTGGTTGATGACGACTAGTGGTTCATCTGGCACGCC 518
 |||
 Qy 101 LeuArgAsnTTPLeuGlyLeuProlIleAspIleArgAsnValCysValLeuPheAlaPro 120
 |||
 Db 519 CTGGCGAACCTGGTTGGGCTTGGCCATTTGACATCAGAAACGTTTGTGTGTTATTTGGGCA 578
 |||
 Qy 121 LeuPheSerGlyValThrAlaTTPAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 |||
 Db 579 CTATTTTCTGGGGTCACCGCTCGGGGACTTACGAAATTTACGAAAGAGATTAAGATGCC 638
 |||
 Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 |||
 Db 639 AGCGCTGGGCTTTTGGCTGCTGTTTATAGCCATTGTCCCGGTTATATATCTAGATCA 698
 |||
 Qy 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 |||
 Db 699 GTGGCGGGCTCTACGATAATAGGGCCATTGCCATTACACTATTAATGTCCTACTTTCATG 758
 |||
 Qy 181 PheTTPleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 |||
 Db 759 TTTTGGATTAGGCCCAAAAGACTGGCTCTATCATGACGACGCTGTCGACCTTTATTC 818
 |||
 Qy 201 TyrPheTyrMetValSerAlaTTPGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
 |||
 Db 819 TACTTCTACATGGTGTGGCTTGGGGTGGATACGTGTTTCATCACCACCTGATCCCACTC 878
 |||
 Qy 221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTTPThr 240
 |||
 Db 879 CATGCTCTTTTGTGATTTTGTGGCGAGATATCTGCAAACTGATTTCTGCCACACC 938
 |||
 Qy 241 ThrTTPYAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 |||
 Db 939 ACTTGGTACGCTATTGAACTGTTCGATCCATCCAGATCCCATTTGTCGGTTTCCTACT 998
 |||
 Qy 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 |||
 Db 999 ATCAGTCTACAGCACCACTGGCGGCACTGGGGTGTTCGGTTTGTGATTTCAGATTGTCCGC 1058
 |||
 Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 |||
 Db 1059 TTCGGTCACTTCGTGGAAGGCCAAATCAGCACAGCTAAGTTTAAAGTCATCATGATGGTT 1118
 |||
 Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 |||
 Db 1119 TCTCTGTTTGTGATCTTGGTCTTGGTGGTGGACTTTTCGCTTGGCTATATATGGGG 1178
 |||
 Qy 321 LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTTPAspThrAsnTyrAlaLysIle 340
 |||
 Db 1179 TTGATTGCCCTCTGGACTGGTAGATTTTATTCGTTATGGGATACCAACTACCAAGATC 1238
 |||
 Qy 341 HisIleProlIleAlaSerValSerGluHisGlnProValSerTTPProAlaPhePhe 360
 |||
 Db 1239 CACATTCCTATCATGCTCGCTTCGGTTTCGGAACATCAACCGGTTTCGTGGCCCGCTTC 1298
 |||
 Qy 361 PheAspThrHisPheLeuIleTTPLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 |||
 Db 1299 TTTGATACCCCACTTTTGTGATCTGGCTATTCCCGCGCGGTGATTCCTACTATTCCTCGAC 1358
 |||

QY 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
 DB 1359 TTGAAGAGCAGCAGCGTTTGTGATCGCTTACTCCGTTCTGTTGCTGACATTGCGCGT 1418
 QY 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerLalaValAla 420
 DB 1419 GTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1478
 QY 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrValAla 440
 DB 1479 TTGTCAGAGATTTGATCATCTACTGGAATTCAGACAGATGACCGCAATACGCCATC 1538
 QY 441 LysProLalaLeuLeuLysLeuLeuValSerGlySerPheLeuThrLeuThr 460
 DB 1539 AAACCTGCGGCACTACTGCGCAATGATGTTTCCGGAATCATCTATTTATTGAT 1598
 QY 461 LeuPheValPheHisSerThrTyrValThrArgThrAlaTyrSerSerProSerValVal 480
 DB 1599 CTTTTGCTTCCATCTTACTTGGTAAACAGAACTGCATCTCTCTCTCTCTCTCTCT 1658
 QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuLysAspPheArgLysValTyr 500
 DB 1659 TTGCCATCACAAACCCAGATGTAATTTGGCGTTGATCGACACTTCAGGAGCGATC 1718
 QY 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTyrTrpAspTyrGly 520
 DB 1719 TATTGTTAAGAACTGAATCTGATGAGCAGTAAGGTTGCACCGTGGGATACGGT 1778
 QY 521 TyrGlnIleGlyGlyMetAlaAspArgThrLeuValAspAsnSerThrTrpAsn 540
 DB 1779 TACCAAAATGTTGGATGCGCAGACAGAACCATTTAGTCGATAACACAGCGGGAACAT 1838
 QY 541 ThrHisIleAlaIleValGlyLeuAlaMetAlaSerProGluGlySerTyrGluLeu 560
 DB 1839 ACTCATCGGCATCGTTGTTAAAGCCATCGCTTCCCTGAAGAGAAATCTTACCAAA 1898
 QY 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
 DB 1899 CTAAAGAGCATGATGTCGATTAATGTTGTCATCTTGGTGGTCTAATTTGGGTTGGT 1958
 QY 581 GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu 600
 DB 1959 GGTGATGATCATCAAAATCTTGTGATGATGATGATGATGATGATGATGATGATGAT 2018
 QY 601 GluIleLysGluArgTyrPheTyrThrAlaGluGlyLysTyrArgValAspAlaArgAla 620
 DB 2019 GAGATAAAGAGCGTTATTTCTATCCGACAGAGGAGATACAGATAGATCAAGGCT 2078
 QY 621 SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu 640
 DB 2079 TCTGAGCATGAGGAATCTGCTACTTTAAGATGCTCCACAAATGATCCACAAATTA 2138
 QY 641 PheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAspVal 660
 DB 2139 TTCAATGTCGCGCAAGCCATGACAGATGCGTCAACAATGATCACACATTAGAGCTC 2198
 QY 661 ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgLysTyr 680
 DB 2199 CCACCATTAGACTACTTTCGAGCAAGTTTACTTCCGAAACACTGGAATGATTAGATAT 2258
 QY 681 GlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
 DB 2259 CAATTGAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2318
 QY 701 SerSerThrLysThrArgArgSerLysArgProGluLeuGlyLeuArgVal 718
 DB 2319 TCTTCTAGAAACACAGAGGTCCTAATAGAGACCTGAATAGGCTTGAGATC 2378
 RESULT 3
 SCYGL022W 3633 bp DNA linear PLN 11-AUG-1997
 LOCUS
 DEFINITION S.cerevisiae chromosome VII reading frame ORF YGL022w.

ACCESSION 272544 Y13135
 VERSION 272544.1 GI:1322488
 KEYWORDS
 SOURCE
 ORGANISM
 Saccharomyces cerevisiae (baker's yeast)
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 REFERENCE 1 (bases 1 to 3633)
 AUTHORS Hebling,U., Hofmann,B. and Delius,H.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3633)
 AUTHORS MIPS
 TITLE Direct Submission
 JOURNAL Submitted (14-MAY-1996) Data collected by MIPS on behalf of the
 European yeast chromosome VII sequencing project. MIPS at the
 Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
 Martinsried, FRG; E-mail: Mews@mips.emblnet.org
 FEATURES
 Location/Qualifiers
 1..3633
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /db_xref="taxon:4932"
 /chromosome="VII"
 complement(1..953)
 /note="unnamed protein product; ORF YGL022c"
 /codon_start=1
 /protein_id="CAA96723.1"
 /db_xref="GI:1945290"
 /db_xref="GOA:P53191"
 /db_xref="Swiss-Prot:P53191"
 /translation="MTALHSVSKTPAIKEBEDGDEDRGVPGLGPRNHDYGRKGD
 ESGATVSPITFEKKIAPRASHSEQSILSSILKMNWQHQQLQOQESGAT
 GVDKQIQIPAMVILKNSAPENVRSKSLGEGDIDGRKASAKSEIKTUPFT
 DDQSNPELDTNSVDVGRKTKTKTNELEDDDDDDVQRKLTQALRLKLS
 FRKASNLRLSKENKAKSSSTSSVSSSTSKVENIVDKLTITSSSSMSQLRFGN
 TNVILDSVNHAAKPPHQMLRKPSELPQASSTNLNFS"
 1253..3409
 /gene="STT3"
 1253..3409
 /gene="STT3"
 /note="ORF YGL022w"
 /codon_start=1
 /protein_id="CAA96722.1"
 /db_xref="GI:1322488"
 /db_xref="GOA:P39007"
 /db_xref="SGD:S0002990"
 /db_xref="Swiss-Prot:P39007"
 /translation="KGSRSCLSVSPQTILKLVIALGAAISSELFVAVIKFETIH
 EFDWFTFRATKILVNNSPFKLNWFDRTWELGRVTGGLYPLGMLTSAFIWALR
 NWLGLDIRNVCLFAPLPSGVTAMATYEFTKEIKDASAGLLAGFIAIVPGIYSR
 VAGSYDNEAIAITLLMTVFWTKAQKTSIHATCAALFYFYMVSAGGYVFTNL
 PLHVLILMGYSKLYSAYTWTYALTVASMQIPFVGLPIRNSNDMAALGVFLI
 QIVAFDGVKQIQISTAKFKVIMVSLFLILVGLVGLSALTVMGLIAPTHVFLV
 NYAKIHPIITASISSEHPQVSPAPFDTHFLIWPAGVFLILDKDEHVFVLA
 VLSVFAAGVRLMLTLTPVCVSAVALSKIPDIYLDKTSRDKVAIPAALLAKLI
 VSGSFIYLVLFVPSHVTAVSPVLPSTQDGLALIDDEREAYVLRMSD
 ESKVAMWDYQIQIGMAORTLVNDNTNTHIALVCKMASPEEKSYELKSHDV
 DVLVYFGLLIGFGGDINDLWIRISIGINPEEKERDFTTAEGEYVDRASETM
 RNSLLKMSYKDFQLENGQATDRVQMITDLPVPLDYDFDEVFTSNWVRIYQL
 KDDAQQGSLTRDVGELTRSTKTRRSIKRPELGRV"

Alignment Scores:
 Pred. No.: 0 Length: 3633
 Score: 631.00 Matches: 717
 Percent Similarity: 99.72% Conservative: 0
 Best Local Similarity: 99.72% Mismatches: 1
 Query Match: 87.88% Indels: 2
 DB: 8 Gaps: 0
 US-10-028-384-6 (1-718) x SCYGL022W (1-3633)
 QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20

[illegible]

Db	2333	TTTGATACCACACTTTTGTGATCTGGCTATTCCCGCGGTGTATTCCTACTATTCTCGAC	2339
Qy	381	LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly	400
Db	2393	TTGAAAGACGACGACGTTTGTTCATCGCTTACTCGCGTTCTGTGTTCGTACTTTGCCGGT	2452
Qy	401	ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla	420
Db	2453	GTTATGGTTAGATTGATGTTGACTTTTGACACACAGTCATCTGTGTCTCGCGCGGTGCGA	2512
Qy	421	LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle	440
Db	2513	TTGTCCAAAGATATTGACATCTACTCGGATTTCACAGCAAGTGACCGCAAAATACGCCATC	2572
Qy	441	LysProAlaIleLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr	460
Db	2573	AAACTCGGCACACTACTGGCCAAATTGATTGTTCCGGATCATTCATCTTTATTGTGAT	2632
Qy	461	LeuPheValPheHisSerThrTTPValThrArgThrAlaTyrSerSerProSerValVal	480
Db	2633	CTTTTCGTCCTCCCATCTTACTTGGGTAAACAAGAACTGCATACCTCTCTCTCTGTGTGT	2692
Qy	481	LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr	500
Db	2693	TTGCCATCACAAACCCAGATGGTAAATTGGCGTTGATCGACGACTTCAGGGAAAGCGTAC	2752
Qy	501	TyrTTPLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTTPTrpAspTyrGly	520
Db	2753	TATTGGTTAAGAATGAACCTCTGATGAGGACAGTAAGGTTGACAGCGTGTGGGATACCGT	2812
Qy	521	TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn	540
Db	2813	TACCAAAATTGGTGGCATGCGACAGACAGAACCATTTAGTCGATAACAACACGTGGAAACAAT	2872
Qy	541	ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle	560
Db	2873	ACTCATCGCGCATCTGTGTAAGACCATGGCTTCCCTCGAAGAGAAATCTTACGAAAT	2932
Qy	561	LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly	580
Db	2933	CTAAAGAGCATGATGTCGATTATGTCCTGGTCATCTTTGGTGGTCTAATTGGGTTTGGT	2992
Qy	581	GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu	600
Db	2993	GGTGATGACATCAACAAATCTCTGTGATGATCAGAAATTAGCGAGGGAATCTGGCCAGAA	3052
Qy	601	GluIleLysGluArgTyr - PheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAl	620
Db	3053	GAGATAAAGAGCGTGA - TTTCTATACCGCAGAGGAGGAATAACAGAGTAGATGCAAGGC	3111
Qy	620	asrGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLe	640
Db	3112	TTCTGAGACCAAGAGAACTCGCTACTTTACAGATGTCCTCAAGAAGATTTCCTCAAAAT	3171
Qy	640	uPheAngGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa	660
Db	3172	ATTCAATGGTGGCCCAAGCCACTGACAGAGTGGCTCAACAAATGATCACACCATTAGACGT	3231
Qy	660	lProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyr	680
Db	3232	CCACCATTTAGACTACTTTCGACGAAGTTTTTTACTTCCGAAAACTGGATGGTTAGGAATAA	3291
Qy	680	rGlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrAr	700
Db	3292	TCAATTGAAGAAGGATGATGCCCAAGGTAGAACTTTGAGGAGCGTTGGTGAGTTAAACAG	3351
Qy	700	gSerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal	718
Db	3352	GTCTTCTACGAAACCAAGGTCCTAATAAGACACCTGAAATTAGGCTTGAGAGTC	3406
RESULT 4			
AX595714		2000 bp	DNA
LOCUS	AX595714		linear
			PAT 14-FEB-2003

```

DEFINITION Sequence 1368 from Patent EP1258494.
ACCESSION AX595714
VERSION AX595714.1 GI:28396965
KEYWORDS Saccharomyces cerevisiae (baker's yeast)
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
REFERENCE 1
AUTHORS Bauer, A., Gavin, A. C., Grandi, P., Krause, R., Kruse, U., Kuster, B.,
Mariöch, M., Schult, J., and Superti-Furga, G.
TITLE Multiprotein complexes from eukaryotes
JOURNAL Patent: EP 1258494-A 1368 20-NOV-2002;
CELLZONE AG (DE)
FEATURES
    source
        1..2000
            Location/Qualifiers
                /organism="Saccharomyces cerevisiae"
                /mol_type="unassigned DNA"
                /db_xref="taxon:4932"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2000
Score: 605.00 Matches: 665
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 1
Query Match: 84.26% Indels: 2
DB: 6 Gaps: 0

US-10-028-384-6 (1-718) x AX595714 (1-2000)
QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
DB 1 ATGGGATCCGACCGGCTGCTGTTTCTGTTGTTTTCAGACCATCTCAAGCTCGTCATC 60
QY 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaIleLysPheGlu 40
DB 61 TTCGTGGCGGATTTTGGGGCTGCCATATCATACAGTTTGTTCAGCATCACTCAAAATTGAG 120
QY 41 SerIleIleHisGluPheAspProTTPheAsnTTPheArgAlaThrLysTyrLeuValAsn 60
DB 121 TCTATTATCCATGAATTCGACCCCTGGTTCAATTATAGGCTACCAAAATATCTCGTCAAC 180
QY 61 AsnSerPheTyrLysPheLeuAsnTTPheAsnTTPheAspArgTTPyrProLeuGlyArg 80
DB 181 AATTCGTTTTCAGATTTTGAAGCTGTTTTCAGACCGTACCTGGTACCCCTCGGAGG 240
QY 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTTPHisala 100
DB 241 GTTACTGGAGGAGCTTTTATATCTCTGGTTTTCATGACGACTAGTGGCTTCATCTGGCAGCC 300
QY 101 LeuArgAsnTTPLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
DB 301 CTGGCCAACTGGTTGGGCTTGGCCATTTGACATCAGAAAGCTTGTGTGTATTTGGCCA 360
QY 121 LeuPheSerGlyValThrAlaTTPAlaThrTyrGluPheThrLysGluIleLysAspAla 140
DB 361 CTATTTCCTGGGGTCCCGCTGGGCGACTTACGAAATTTACGAAAGAGATTAAAGATGCC 420
QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
DB 421 AGCGCTGGGGCTTTGGCTGTGGTTTATAGCCATTTGTCCTGGTTATATATCTAGATCA 480
QY 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
DB 481 GTGGCGGGGCTCTACGATAATAGGCCATTTGCCATTTACACTATTATGTGTCACTTCATG 540
QY 181 PheTTPileLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
DB 541 TTTTGGATTAAAGGCCAAAAGAGCTGGCTCTATCATGACGCAACGTCGTCAGCTTTATTC 600
QY 201 TyrPheTyrMetValSerAlaTTPGlyTyrValPheIleThrAsnLeuIleProLeu 220
DB 1740
601 TACTTCTACATCGTGTGGCTGGGTGATACGTTGTTCAACCAACTTGATCCCACTC 660
221 HisValPheLeuLeuLeuMetGlyArgTyrSerLysLeuTyrSerAlaTyrThr 240
661 CATGTCTTTTCTGATTTTGTATGGCAGATATTCGTCACCAACTGTATTTCTGCTACAC 720
241 ThrTTPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
721 ACTTGGTACGCTATTGGAACTGTTGCATCCATGAGATCCCATTTGTCGGTTCTTCACT 780
261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
781 ATCAGGTCTTAACGACCATCGCGCATTTGGGTGTTTTCGGTTTGTATTCAGATTGTCGC 840
281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
841 TTCGGTGAATTCGTGAGGGCCAAATCAGCACAGCTAAGTTTAAAGTCATCATGATGGTT 900
301 SerLeuPheLeuLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
901 TCTCTGTTTTCATCTGCTGCTGTTGGTGGTGGTCTTTCGCTTTCAGCTATATGGGG 960
321 LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTTPAspThrAsnTyrAlaLysIle 340
961 TTGATTGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCAACTACGCAAGATC 1020
341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTTPProAlaPhePhe 360
1021 CACATTCCTATCATTCGCTCGCTTTCGAAACATCAACCCGTTTCGGTGGCCGCTTTCCTC 1080
361 PheAspThrHisPheLeuIleTTPLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
1081 TTTGATACCACTTTTTCATCTGGCTATTCGCCCGCGGTGTATTCCTACTATTCTCTGAC 1140
381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
1141 TTGAAGACGACGACGCTTTTGTTCATCGCTTACCTGCTTCTGTGTTCTGTTTTCGGGT 1200
401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
1201 GTTATGGTTAGATTGATGTTGACTTTGACACCACTCATCTGTGTGTCGCGCGCGTCGCA 1260
421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
1261 TTGTCGAAAGATTTTGACATCTCTGATTTTCAAGACAGTACGCGCAAAATACGCCATC 1320
441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
1321 AAACCTGGCGCACTACTGGCCAAATTTGATTTGTTTCGGATCATTCATCTTTTATTTGTAT 1380
461 LeuPheValPheHisSerThrTTPValThrArgThrAlaTyrSerSerProSerValVal 480
1381 CTTTTCGTTCTTCATCTCTACTTGGGTAAACAGAACTGACACTCTCTCTCTCTCTGTTGT 1440
481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
1441 TTGCCATCACAAACCCAGATGGTAAATTTGGCGTTGATTCGACGACTTCAGGGAAGCGTAC 1500
501 TyrTTPLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTTPThrPheTyrGly 520
1501 TATTGGTTAAGATGAATCTCTGATGAGGACAGTAAGGTTGACGCTGGTGGGATTTACGT 1560
521 TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
1561 TACCAAAATTTGGTCATGGCAGACAGAACCCATTTAGTTCGATTAACCAACAGCTGGAAAT 1620
541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle 560
1621 ACTCATCATCGCATCTGTTGGTAAAGCCATGGCTTCCTCCCTGAGAGAAATCTTACGAAAT 1680
561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyLeuIleGlyPheGly 580
1681 CTAAAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740

```


Db 1321 AAACCTGGCGCACTACTGGCAAAATGATTGTTTCGGATCAATTCATCTTTTATTGTAT 1380
 Qy 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTrpSerProSerValVal 480
 Db 1381 CTTTTCGCTTCCATCTACTTGGGTACAGAACTGCATACCTCTCTCTCTGTGT 1440
 Qy 481 LeuProSerGlnThrProAspGlyLeuAlaLeuLeuLeuLeuLeuLeuLeuLeuLeu 500
 Db 1441 TTGCCATACAAACCCAGATGTAATTTGGCTTGATCGACGACTTCAGGGAAGCGTAC 1500
 Qy 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTrpGly 520
 Db 1501 TATTGGTTAAGATGATCTCTGATGAGACGATGAGGTTCCACGCTGGTGGATACGGT 1560
 Qy 521 TyrGlnLeuGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
 Db 1561 TACCAAAATGGTGGCATGGCAGACAGAACCACTTTAGTCGATTAACACACGTTGAACAT 1620
 Qy 541 ThrHisLeuAlaLeuValGlyLeuAlaMetAlaSerProGluGluLysSerTrpGluLeu 560
 Db 1621 ACTCATCTGGCATGCTTGGTAAAGCATGGCTTCCCTGACAGAAATCTACGAAAT 1680
 Qy 561 LeuLysGluHisAspValAspTyrValLeuValLeuPheGlyLeuLeuGlyPheGly 580
 Db 1681 CTAAAGAGCATGATGTCGATATGCTTGGTCATCTTGGTGTCTTAATGGGTGGT 1740
 Qy 581 GlyAspAspLeuAsnLysPheLeuTrpMetLeuArgLeuSerGluGlyLeuTrpProGlu 600
 Db 1741 GGTGATGATCATCAACAAATCTTGGTATGATGATGATGATGATGATGATGATGATG 1800
 Qy 601 GluLeuLysGluArgTyr-PheTyrThrAlaGluGlyGluTyrArgValAlaAlaArgAl 620
 Db 1801 GAGATAAAGAGCGTGA-TTCTATACCGCAGAGGGAGATACAGAGTAGATGCAAGGC 1859
 Qy 620 aSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLe 640
 Db 1860 TTCTGAGACCATGAGGAATCGCTACTTTACAGATGTCCTACAAAGATTTCCCAAT 1919
 Qy 640 uPheAsnGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa 660
 Db 1920 ATTCAATGGTGGCAAGCCACTGACAGAGTGGCTGCAAAATGATCACCACTTAGCGT 1979
 Qy 660 lProProLeuAspTyrPhe 666
 Db 1980 CCCACCATAGACTACTTC 1998

RESULT 6

AX830928 2000 bp DNA linear PAT 12-DEC-2003
 LOCUS Sequence 1648 from Patent WO03072602.
 DEFINITION AX830928
 ACCESSION AX830928
 VERSION 1 GI:39839809
 KEYWORDS Saccharomyces cerevisiae (baker's yeast)
 SOURCE Saccharomyces cerevisiae
 ORGANISM Saccharomyces cerevisiae

REFERENCE 1
 AUTHORS Baker, A., Gavin, A.C., Superti-Furga, G., Kuester, B., Schultz, J.,
 Marzioch, M., Grandi, P., Krause, R., Kruse, U., Merino, A., Bauch, A.,
 Michon, A.M., Leutwein, C. and Rick, J.

TITLE Protein complexes and methods for their use
 JOURNAL Patent: WO 03072602-A 1648 04-SEP-2003;
 CELLZONE AG (DE)
 FEATURES location/Qualifiers
 1..2000
 /organism="Saccharomyces cerevisiae"
 /mol_type="unassigned DNA"
 /db_xref="taxon:4932"

ORIGIN

Alignment Scores:
 Pred. No.: 0 Length: 2000

Score: 605.00 Matches: 665
 Percent Similarity: 99.70% Conservative: 0
 Best Local Similarity: 99.70% Mismatches: 1
 Query Match: 84.26% Indels: 2
 DB: Gaps: 0

US-10-028-384-6 (1-718) x AX830928 (1-2000)

Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 Db 1 ATGGGATCCGACCGGTCGTGTTGTTGCTGTGTTTTCAGACCATCTCTCAAGCTCGTCATC 60
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 61 TTGCTGGCGAATTTTGGGGCTGCCATATCATCAGCTTGTGTTGCGAGTCATCAATTTGAG 120
 Qy 41 SerIleHisGluPheAspProTrpPheAsnTyrArgAlaThrLysTyrLeuValAsn 60
 Db 121 TCTATTATCATGAATTCGACCCCTGGTTCATATTATAGGGCTACCAATATCTCGTCAAC 180
 Qy 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArg 80
 Db 181 AATTCGTTTTCAGAGTTTGAACCTGTTTGCACCGCTACCTGTTACCCCTCGGAAGG 240
 Qy 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrPheAla 100
 Db 241 GTTACTGGAGGACATTTATCTGCTGTTTGTATGACGACTAGTGGCTTCATCTGGCAGCC 300
 Qy 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgLeuValCysValLeuPheAlaPro 120
 Db 301 CTGCGCAACTGTTGGGCTTGGCCATTGACATCAGAAACGTTTGTGTGCTATTTCGCGCA 360
 Qy 121 LeuPheSerGlyValThrAlaThrAlaThrTyrGluPheThrLysGluLeuLysAspAla 140
 Db 361 CTATTTCCTGGGCTCACCGCTGGCGACTTACGAATTTACGAAGAGATTAAGATGCC 420
 Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 Db 421 AGGCTGGGCTTTTGGCTGCTGTTTATAGCAATTTGCCCGGTTATATCTAGATCA 480
 Qy 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 481 GTGGCGGGTCTCTAGCATAATGAGCCATTGCCATTACACTATTATTAATGGTCACATTTCATG 540
 Qy 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 Db 541 TTTTGGATTAAAGCCCAAAAGACTGGCTCTATCATGACGCAACGTCGTGAGCTTTATTTC 600
 Qy 201 TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
 Db 601 TACTTCTACATGGTGTGGGCTTGGGCTGGATACGTTGTTTCATACCAACTTGTATGCCCATC 660
 Qy 221 HisValPheLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 Db 661 CATGCTTTTCTGCTGATTTTGTATGGGAGATTTCTGCCAAGCTGATTCTGCTACACC 720
 Qy 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 721 ACTGTGATCGCTATTGGAACTGTTGCATCCATGAGATCCCATTTGTCGGTTCCTACCT 780
 Qy 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 781 ATCAGGTCTAACGACCAACATGGCCGCTTGGGTGTTTTCGGTTGATTTCAGATTGTCGCC 840
 Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 Db 841 TTGCTGATCTCGTGAAGGCGCAATTCAGACACAGTAAGTTTAAAGTCATCATGTGTT 900
 Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 Db 901 TCTCTGTTTTCGATCTTGGTCTTGGTGTGGTCTTTCGAGCTTTCTGCTTGCCTATATGGGG 960
 Qy 321 LeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle 340

Db 961 TTGATTGCCCTTGACCTGGTAGATTTTATTCGTATGGGATACCACTACGCAAGATC 1020
 QY 341 HSLLEPRLLEILALASERVALSERGLUHISGLNPROVALSERTRPProAlaPhePhe 360
 Db 1021 CACATTCCTATCATTCCTCGTTTCGAAACATCAACCCGTTTCGTGGCCGCTTTCTTC 1080
 QY 361 PheAspThrHisPheLeuLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 Db 1081 TTTGATACCCACTTTTTCATCTGGCTATTCCCGCCGGTGATTCTCTACTATTCTTCGAC 1140
 QY 381 LeuLysAspGluHisValPheValIleAlaLysSerValLeuLeuCysSerTrpPheAlaGly 400
 Db 1141 TTGAAGACGACGACGCTTTTGTCTATCCCTTACTCCGTTCTGTTCTGTTCTGTTCTGCGGT 1200
 QY 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
 Db 1201 GTTATGGTTAGATTGATGTGACTTTGACACCACTGATCTGTGTGTCGCGCGCTGCGCA 1260
 QY 421 LeuSerLysIlePheAspIleTrpLeuAspPheLysThrSerAspArgLysTrpAlaIle 440
 Db 1261 TTGTCCAGATATTTGACATCTACTGCTGATTTCAAGACAAAGTGACCCCAATACGCCATC 1320
 QY 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTrpLeuTrp 460
 Db 1321 AAACGTGGGACACTACTGCCAATTTGATGTTTCGCGATCATTCATCTTTTATTTGTAT 1380
 QY 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTrpSerSerProSerValVal 480
 Db 1381 CTTTTCGTCTCCATTTACTTGGGTAAACAAGAACTGCATCTCTCTCTCTCTCTCTCTGTT 1440
 QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspPheArgGluAlaTrp 500
 Db 1441 TTGCCATCACAAACCCAGATGGTAAATTTGCGGTGATCGACGACTTCAGGAGCGTAC 1500
 QY 501 TrpTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTrpGly 520
 Db 1501 TATTGGTTAAGAAATGAATCTCTGATGAGACAGTAGTAAAGTTGCGCGTGGTGGATACGGT 1560
 QY 521 TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
 Db 1561 TACCAATTTGGTGGATGGCAGACAGAACCACTTTAGTCGATACACACAGTGGACAT 1620
 QY 541 ThrHisIleAlaIleValGlyValAlaMetAlaSerProGluGluLysSerTrpGluIle 560
 Db 1621 ACTCATCGCCATCGTTGGTAAAGCCATGGCTTCCCTGGAAGAGAAATCTTACGAAAT 1680
 QY 561 LeuLysGluHisAspValAspTrpValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
 Db 1681 CTAAAGAGCATGATGTCATATGCTTGTGTCATCTTGGTGTCTAATGGTTGGT 1740
 QY 581 GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu 600
 Db 1741 GGTGATGATCATCAAAATTTCTGTGGATGATCAGAATTAGCGAGGGAATCTGGCCAGAA 1800
 QY 601 GluIleLysGluArgTyr-PheTrpThrAlaGluGlyGluTrpArgValAspAlaArgAl 620
 Db 1801 GAGATAAAGCGGTGA-TTTCTATACCGAGAGGGAGATACAGATAGATGCAAGGCG 1859
 QY 620 aserGluThrMetArgAsnSerLeuLeuTrpLysMetSerTrpLysAspPheProGlnLe 640
 Db 1860 TTCTGAGACCATGAGCAATCTGCTACTTTTACAAGATGCTCTACAAGATTTCCCAAT 1919
 QY 640 uPheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAspVa 660
 Db 1920 ATTCAATGGTGCCAGCCACTGACAGAGTGGTCAACCAATGATCATCAACATTAACGT 1979
 QY 660 lProProLeuAspTrpPhe 666
 Db 1980 CCCACCATTAGACTACTTC 1998

RESULT 7
 CQ450623

LOCUS CQ450623 381 bp DNA linear PAT 30-JAN-2004
 DEFINITION Sequence 16383 from Patent WO0192523.
 ACCESSION CQ450623
 VERSION CQ450623.1 GI:41419542
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 1
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 Shimkets, R.A. and Leach, M.D.
 Human polynucleotides and polypeptides encoded thereby
 Patent: WO 0192523-A 16383 06-DEC-2001;
 Curagen Corporation (US)
 LOCATION/Qualifiers
 1..381
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 FEATURES
 source
 1..381
 6.63e-122 Length: 381
 121.00 Matches: 121
 100.00% Conservativity: 0
 100.00% Mismatches: 0
 16.85% Indels: 0
 6 Gaps: 0
 Alignment Scores:
 Pred. No.: 6.63e-122 Length: 381
 Score: 121.00 Matches: 121
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 16.85% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-6 (1-718) x CQ450623 (1-381)
 QY 452 SerGlySerPheIlePheTrpLeuTrpLeuPheValPheHisSerThrTrpValThrArg 471
 Db 1 TCCGGATCATTCATCTTTTATTTGATCTTTTGGTCTTCCATCTACTTGGGTAAACAAGA 60
 QY 472 ThrAlaTrpSerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAla 491
 Db 61 ACTGCATACCTCTTCT 511
 QY 492 LeuIleAspAspPheArgGluAlaTrpTrpLeuArgMetAsnSerAspGluAspSer 511
 Db 121 TTGATCGACGACTTCAGGAGAGCGTACTATTGGTTAAGATGAACTCTCTGATGAGGACAT 180
 QY 512 LysValAlaAlaAlaTrpTrpAspTrpGlyTrpGlnIleGlyMetAlaAspArgThrThr 531
 Db 181 AAGTTGTCAGCGTGGTGGGATTACGGTTACCAAAATTTGGTGGCATGGCAGACAGAACCT 240
 QY 532 LeuValAspAsnAsnThrTrpAsnAsnThrHisAlaIleValGlyLysAlaMetAla 551
 Db 241 TTAGTCGATACACACAGTGGACATATCTACATCGCCATCGTTGGTAAAGCCATGGCT 300
 QY 552 SerProGluGluLysSerTrpGluIleLeuLysGluHisAspValAspTrpValLeuVal 571
 Db 301 TCCCTCGAAGAGAAATCTTACGAAATTTCTAAAGAGCATGATCGATTATGTCTTGGTC 360
 QY 572 Ile 572
 Db 361 ATC 363
 RESULT 8
 SCALK1
 LOCUS S.Cerevisiae ALK-1 gene. 3447 bp DNA linear PLN 01-AUG-1995
 DEFINITION S.Cerevisiae ALK-1 gene.
 ACCESSION X87672.1 GI:861109
 VERSION X87672.1
 KEYWORDS ALK-1 gene; DNA damage-responsive protein.
 SOURCE Saccharomyces cerevisiae (baker's yeast)
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 1 (bases 1 to 575)
 Yohida, S. and Anraku, Y.
 Isolation and characterization of a novel gene, STT3, which is

involved in protein glycosylation
 Unpublished
 REFERENCE 2 (bases 2742 to 2976)
 AUTHORS Weinstock, K.
 JOURNAL Unpublished
 REFERENCE 3 (bases 3365 to 3447)
 AUTHORS Bidwal, A.P., Reed, J.C. and Glover, C.V.
 TITLE Cloning and disruption of CKII, the gene encoding the 38 kDa beta subunit of Saccharomyces cerevisiae casein kinase II (CKII). Deletion of CKII regulatory subunits elicits a salt-sensitive phenotype
 J. Biol. Chem.
 REFERENCE 4
 AUTHORS Moen, C., Lindstedt, B.A., Berdal, K.G., Rognes, T. and Seeborg, E.C.
 TITLE A novel DNA damage-response gene from Saccharomyces cerevisiae with homology to protein kinase
 Unpublished
 REFERENCE 5 (bases 1 to 3447)
 AUTHORS Rognes, T.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAY-1995) T. Rognes, University of Oslo, Dept of Molecular Biology, Inst. of Medical Microbiology, The National Hospital, 0027 Oslo, NORWAY
 COMMENT Related sequence: Z16294.
 FEATURES
 source
 1..3447
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /db_xref="taxon:4932"
 442..2721
 /gene="ALK-1"
 442..2721
 /gene="ALK-1"
 /codon_start=1
 /protein_id="CAA61012.1"
 /db_xref="GI:861110"
 /db_xref="GOA:P43633"
 /db_xref="SGD:S0002989"
 /db_xref="Swiss-Prot:P43633"
 /translation="MLHFDIVQLLSKSHQVPPMPDTSFEFVDDKRFIAL
 EVSDNDDCDTADTADLESALSKWRESADLNVTGNTSKRTSSKKRWSL
 LSNHSAVSSKSKRWSLVSSFTSESHKDRSNVLOQRKLSQSSLDTVASNS
 ISASRLKSTGLSLRQLFTGINDISQFGIGIPQKKNLPPTMGKNSASTS
 SENLRPLPLVNHKRPPTQPOQPLNYSASLRSSISSTVSSSSSKWRFWR
 NKQTPALLOPDHSLKTPAVNRDSMTPEPRNVKHTFSFDPKTIFFSNITYSE
 SSDTSSMETLTKASSSLSLNLVKENSOSLKHKSSHASLOKFRNKGKSMIA
 PSVATNSNDDSCSYSKNSTLSHRISLPVPOVSRDKLQNLRYSTLSLNSKSL
 PMKNDHDELLQLLNCIDIKILNPAKGVDPPLINDVNHLSICQLTSNVWQIGV
 CKVSLGTIDITWDRKFLSLQLEKLMQKQFGIPOLLSKSFVKEANGGLYLILL
 FKDHGTPISLNKKQILKIPWSCAGIHGLEKNLKFHRLNLDNLIDGNNT
 IIDFKCRLQTPQDDVLYLRDLHPLFLNGKDKSKINERYQVQFEPEIYQSVRLNMD
 ASAFEPNLYLWYLSRVLLKFGDKLGNKDNRDKQARVINHLNMLAVHKGGL
 FKRLTEIDIKNTGOLLKAL"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,22e-66 Length: 3447
 Score: 71.00 Matches: 71
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.89% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x SCALK1 (1-3447)

QY 648 AspArgValArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyrPheAsp 667
 DB 2 GACAGAGTGGCTGCAACAAATGATCACACCATAGAGTCCACCATAGACTTACGAC 61
 QY 668 GluValPheThrSerGluAsnTyrMetValArgIleTyrGlnLeuLysAspAspAla 687
 DB 62 GAAGTTTACTTCCGAAGAACTGGAGTGTAGATATATCAATTGAGAGGATGATGCC 121

QY 688 GlnGlyArgThrLeuArgAspValGlyGlnLeuThrArgSerSerThrLysThrArgArg 707
 DB 122 CAAGGTAGAACTTTGAGGGACGTTGGTGAGTTAACCAAGGCTTCTACGAAACCAAGG 181
 QY 708 SerIleLeuArgProGlnLeuGlyLeuArgVal 718
 DB 182 TCATTAAGAGACCTGAATTAGGCTTGAGATC 214

RESULT 9
 CR382123_09/c
 WPCOMMENT
 Sequence split into 18 fragments LOCUS CR382123 Accession CR382123

Fragment Name	Begin	End
CR382123_00	1	110000
CR382123_01	100001	210000
CR382123_02	200001	310000
CR382123_03	300001	410000
CR382123_04	400001	510000
CR382123_05	500001	610000
CR382123_06	600001	710000
CR382123_07	700001	810000
CR382123_08	800001	910000
CR382123_09	900001	1010000
CR382123_10	1000001	1110000
CR382123_11	1100001	1210000
CR382123_12	1200001	1310000
CR382123_13	1300001	1410000
CR382123_14	1400001	1510000
CR382123_15	1500001	1610000
CR382123_16	1600001	1710000
CR382123_17	1700001	1753957

Continuation (10 of 18) of CR382123 from base 900001 (CR382123 Kluveromyces lactis str)

Alignment Scores:
 Pred. No.: 4.86e-30 Length: 110000
 Score: 39.00 Matches: 39
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x CR382123_09 (1-110000)

QY 512 LysValAlaAlaTyrTyrAspTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThr 531
 DB 103607 AAAGTTGCAGCTTGTGGGATATGTTATCAATCGGTGCTGATAGAACCACT 103548

QY 532 LeuValAspAsnAsnThrTyrAsnAsnThrHisIleAlaIleValGlyLysAlaMet 550
 DB 103547 CTAGTTGATTAATACACTTGAATAACACCCACCATTCATTGTTGTAAGGCAATG 103491

RESULT 10
 CR382123_10/c
 WPCOMMENT
 Sequence split into 18 fragments LOCUS CR382123 Accession CR382123

Fragment Name	Begin	End
CR382123_00	1	110000
CR382123_01	100001	210000
CR382123_02	200001	310000
CR382123_03	300001	410000
CR382123_04	400001	510000
CR382123_05	500001	610000
CR382123_06	600001	710000
CR382123_07	700001	810000
CR382123_08	800001	910000
CR382123_09	900001	1010000
CR382123_10	1000001	1110000
CR382123_11	1100001	1210000
CR382123_12	1200001	1310000
CR382123_13	1300001	1410000
CR382123_14	1400001	1510000
CR382123_15	1500001	1610000
CR382123_16	1600001	1710000

```

CR382123 17 170001 1753957
Continuation (11 of 18) of CR382123 from base 1000001 (CR382123 Kluyveromyces lactis str
Alignment Scores:
Pred. No.: 4,866-30 Length: 110000
Score: 39,00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x CR382123_10 (1-110000)
QY 512 LysValAlaAlaTrpAspTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThr 531
DB 3607 AAAGTTGACAGCTTGGTGGATATGTTATCAATCGGTGATGGTGTATAGACCACT 3548
QY 532 LeuValAspAsnThrTrpAsnAsnThrHisLeAlaLeValGlyLysAlaMet 550
DB 3547 CTAGTTGATATACACTTGGATTAACACCCACATTCATTGTTGTAAGGCAATG 3491

RESULT 11
AE016903 300029 bp DNA linear PLN 14-APR-2004
LOCUS Erethothecium gossypii chromosome VI, section 5 of 6 of the complete
DEFINITION Erethothecium gossypii (Ashbya gossypii)
ACCESSION AE016903 AE016819
VERSION AE016903.1 GI:44985021
KEYWORDS Erethothecium gossypii (Ashbya gossypii)
SOURCE Erethothecium gossypii
ORGANISM Erethothecium gossypii
REFERENCE 1. (bases 1 to 300029)
AUTHORS Dietrich, F.S., Voegel, S., Brachet, S., Lerch, A., Gates, K.,
Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., Wing, R.A.,
Flavler, A., Gaffney, T.D. and Philippsen, P.
TITLE The Ashbya gossypii genome as a tool for mapping the ancient
Saccharomyces cerevisiae genome
JOURNAL Science 304 (5668), 304-307 (2004)
PUBMED 15001715
REFERENCE 2. (bases 1 to 300029)
AUTHORS Brachet, S., Voegel, S.E., Dietrich, F.S., Lerch, A., Gaffney, T. and
Philippsen, P.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-2002) Applied Microbiology, Biozentrum,
Universitaet Basel, Klingelbergstrasse 50-70, Basel CH-4056,
Switzerland
COMMENT This is low coverage sequence generated to identify the complete
set of genes and the gene order on this chromosome. Regions of low
quality are identified. Before doing extensive work on any gene
identified here the sequence should be confirmed.
FEATURES
Location/Qualifiers
1. 300029
/molecule="Erethothecium gossypii"
/mol_type="genomic DNA"
/strain="ATCC 10895"
/db_xref="arcc:10895"
/db_xref="taxon:33169"
/chromosome="VI"
/note="synonym: Ashbya gossypii"
1. 36701
/molecule="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2404"
26449..57449
/molecule="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag1650"
51501..123305
/molecule="Erethothecium gossypii"

```

```

/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2318"
100002..157868
/molecule="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2457"
130220..181064
/molecule="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2161"
157733..244785
/molecule="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2545"
244643..300029
/molecule="Erethothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2180"
complement(142..2918)
locus_tag="AFR429C"
complement(142..2918)
locus_tag="AFR429C"
product="AFR429Cp"
complement(142..918)
locus_tag="AFR429C"
/note="synthetic homolog of Saccharomyces cerevisiae
YKL207W"
evidence=not_experimental
product="AFR429Cp"
protein_id="AA853800.1"
/db_xref="GI:44985022"
translation="MSLGACACAGADTLDSRLKLVLLPISVVMILGVVRYLMILL
TSKYSVPRAVTENQYISKAIAMLTNGSLHEEAPRMEQVLAGVLSGKYLAKGK
AEPQNVLAADPNTEALMGAKNLAGYIPQVIMWVNYFAGFVLMKLPPLNRF
KEMQSGVTADLADRWSSISWIFSMFGLDPVYVNLJFGDEAASQSMQOMFVGGG
VPGFTPEALMKYANDLTIAHSCDFGIEDRVLMQYAA"
1. 300029
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2180"
complement(142..2918)
locus_tag="AFR429C"
complement(142..2918)
locus_tag="AFR429C"
product="AFR429Cp"
complement(142..2918)
locus_tag="AFR429C"
/note="synthetic homolog of Saccharomyces cerevisiae"
evidence=not_experimental
product="AFR429Cp"
protein_id="AA853801.1"
/db_xref="GI:44985023"
translation="MCRRLDAGHGGAGAFASVPTTQKSVDECPNYGLYAKVPHY
RNGAKTPLNLPYQRPKCRFTSASVDKIFEFVALLKDPDLALFENTFENTLDT
TILWHTVEESRELEQASEIAGMPGRHRLAKQKQAFETFIIVTGDIAEWLSDSA
WQLSYQRFIKSDPFALETKGAIQYAEYKGNPYCNAPHPHSGVGRSSIDDV
LPLNWEQVECKEYDLSASFLTSRQYHENAPDSDRFDFTEDWLAAYKIMKLVK
AQMSFTDTGHPQVYTKFORSTKIASETLPLAGTGNVNTGLVRAFRPDDAT
IROPFVGNVIAVELEHLTAIQSVLGDAAAAPARKVSPSTDLGLLEASQLS
IRADGIMEHVALEHKEGEVAYEYDVGSHLPMDDANIPLLSLPGLGFGVKDDAV
YRTRKMLTEGNYLNGEHFGQGGPHGLHNAPMSLLVRLTSDSDDEIMHCL
DMWNTKGTGLLHESVRAPHPDGLSYTRSNFANSEFAKTLIDAEKRPHLINQ
DAATKPLETFLSQSASV"
1193
/note="region of low quality"
1338
/note="region of low quality"
2809
/note="region of low quality"
2881..2882

```


LOCUS	SPEC1271	S.pombe chromosome II	40907 bp	DNA	linear	PLN 23-JUN-2003
DEFINITION	S.pombe chromosome II cosmid c1271.					
ACCESSION	AL034353					
VERSION	AL034353.2	GI:6912224				
KEYWORDS	Schizosaccharomycetes pombe (fission yeast)					
SOURCE	Schizosaccharomycetes pombe					
ORGANISM	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.					
REFERENCE	1 (bases 1 to 40907)					
AUTHORS	Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brookes, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltham, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E.J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Moore, P., Moulé, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neill, S., Pearson, D., Quail, M.A., Rabinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R.G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volkart, G., Aert, R., Robben, J., Grymonprez, B., Weijens, I., Vanstraelen, E., Rieger, M., Schaefer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritsch, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehmach, H., Reinhardt, R., Pohl, T.M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S.J., Xiang, Z., Hunt, C., Moore, K., Hurst, S.M., Lucas, M., Rochet, M., Galliard, C., Tallada, V.A., Garzon, A., Thode, G., Daga, R.R., Cruzado, L., Jimenez, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J.L., Moreno, S., Armstrong, J., Forsburg, S.L., Cerutti, L., Lowe, T., McComb, W.R., Paulsen, I., Potashkin, J., Shpakovski, G.V., Ussery, D., Barrett, B.G. and Nurse, P.					
TITLE	The genome sequence of Schizosaccharomycetes pombe					
JOURNAL	Nature 415 (6874), 871-880 (2002)					
MEDLINE	21848401					
PUBLISHED	11859360					
REFERENCE	2 (bases 1 to 40907)					
AUTHORS	Lyne, M., Rajandream, M.A., Barrell, B.G., Beck, A., Borzym, K., Klages, S., Langer, I. and Reinhardt, R.					
TITLE	Direct Submission					
JOURNAL	Submitted (23-NOV-1998) European Schizosaccharomycetes genome sequencing project, Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: pombe@sanger.ac.uk and Max Planck Institut fuer Molekulare Genetik, Ihnestr. 73, D-14195 Berlin, Germany					
COMMENT	On Feb 7, 2000 this sequence version replaced gi:3925769.					
Notes:	Details of S. pombe sequencing at the Sanger Institute are available on the World Wide Web (URL, http://www.sanger.ac.uk/projects/s_pombe/). CDS are numbered using the following system eg SPAC5H10.01c. SP (S. pombe), A (chromosome 1), cSH10 (cosmid name), .01 (first CDS), c (complementary strand). However, clones may have been reorientated since the original submission, therefore the complementary strand notation may be invalid for strand inference. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once.					
FEATURES	Location/Qualifiers					
source	1. .40907					
	/organism="Schizosaccharomycetes pombe"					
	/mol_type="genomic DNA"					
	/strain="972h"					
	/db_xref="taxon:4896"					
	/chromosome="II"					
	/map="III"					
	/clone="cosmid c1271"					
	1307. .3367					
gene						

ORGANISM	Schizosaccharomycetes pombe
REFERENCE	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.
AUTHORS	1 (sites)
TITLE	Yoshida, S., Matsura, A., Merregaert, J. and Anraku, Y.
JOURNAL	Schizosaccharomycetes pombe att3+ is a functional homologue of Saccharomyces cerevisiae ST3 which regulates oligosaccharyltransferase activity
MEDLINE	99251095
PUBLISHED	10234787
REFERENCE	2 (bases 1 to 5048)
AUTHORS	Yoshida, S.
TITLE	Submitted (04-JUN-1998) Satoehi Yoshida, Kirin Brewery Co., Ltd., Central Laboratories for Key Technology; 1-13-5, Fukuura Kanazawa-Ku, Yokohama, Kanagawa 236-0004, Japan (E-mail: satoehi@kirin.co.jp, Tel:81-45-788-7360, Fax:81-45-788-4042)
JOURNAL	Location/Qualifiers
FEATURES	1. .5048
source	/organism="Schizosaccharomycetes pombe"
	/mol_type="genomic DNA"
	/db_xref="taxon:4896"
gene	join(1203..1302,1362..3511)
CDS	/gene="att3+"
	join(1203..1302,1362..3511)
	/gene="att3+"
	/note="similar to S. cerevisiae, mouse and human homologue"
	/codon_start=1
	/product="oligosaccharyltransferase subunit"
	/protein_id="BAA96419.1"
	/db_xref="GI:4586803"
	/translation="MANSATITSLKGVKSHQKDWKLPKLVILICIAVASVSRFSVIRYSIIHEPDPWFNPKSLIKVQGFYNLFNWFNDFSRVPLGRVAGTLYPLQAMVTSIFIKFVLLRINYNIDVCLLAPAFSGITATYTLARELKSACGGLAAAFVAGIA PGYTRSVAGSYDEAATITLLMTFALIKAVKSGSSFMGACTGLLYFYMTAMWGIVFTNMIPLHVVYVLMGRVTSKLYIATYVYVIGTASMQVFFVGFQFVSTSEMSALGVGLGLQPLAFYVNVKLVSSKQFQILIRFALVCLVGLATVILFALSTGVIAPTWG RPYSLMDNVAKIHIPIIAVSVEHQPETMSLSPFDLQELIKLIPGVYILCFKELNEH VFIIIVPLGVYFCGVWVRVLUITPCVCARAAVALSTLLDTVMGEVEDSKVSEEA SAKSKKRGISLISLFTSGSKNIGIYSLLSRVLSISSTAIFLIMFVHSSVWTSNAY SSPTVLTSLNDGSLMYIDDREAYDMLRNTPTPKNSWWDYQIAGMDRITL VDNNTWNTHTATIGKAMSPFEKAPYILRKHQVDYILYIGTLYGSDDMKXILM IRISQGLWFEIVERNFTPNGEYRTDAAATPMBRSLLYKMSYHGAWKILFFPNQYD RAROKLPKSDPQPLFTIEAFTVHHLVRLYKVKKPDITLGRDLKQVTLFEEGKRKESA VLOKLTREL"
intron	1303..1361
	/gene="att3+"
ORIGIN	
Alignment Scores:	
Pred. No.:	1.35e-23
Score:	Length: 5048
Percent Similarity:	32.00
Best Local Similarity:	Matches: 32
Query Match:	Conservative: 0
DB:	Mismatches: 0
	Indels: 0
	Gaps: 0
US-10-028-384-6 (1-718) x AB015232 (1-5048)	
QY	322 ileAlaProThrGlyArgPheTy-SerLeuTrpAspThrAsnTYrAlaLysileHis 341
Db	2246 ATCGCTCCTTGACAGGACGGTTCATTCTCTTTGGATACAAACACGCGCAGATTTCAT 2305
QY	342 lleProIleIleAlaSer-ValSer-GluHisGlnPro 353
Db	2306 ATTCTCATCATTCGCTTCGGTATCAGAACATCAGCCT 2341
RESULT 14	
SPBC1271/c	

ORGANISM	Schizosaccharomycetes pombe Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.			
REFERENCE	1 (sites)			
AUTHORS	Yoshida, S., Matsumura, A., Merregaert, J., and Anraku, Y.			
TITLE	Schizosaccharomycetes pombe att3+ is a functional homologue of Saccharomyces cerevisiae ST3 which regulates oligosaccharyltransferase activity			
JOURNAL	Yeast 15 (6), 497-505 (1999)			
MEDLINE	99251095			
PUBLISHED	10234787			
REFERENCE	2 (bases 1 to 5048)			
AUTHORS	Yoshida, S.			
TITLE	Direct Submission			
JOURNAL	Submitted (04-JUN-1998) Satoshi Yoshida, Kirin Brewery Co., Ltd., Central Laboratories for Key Technology; 1-13-5, Fukura Kanazawa-ku, Yokohama, Kanagawa 236-0004, Japan (E-mail: sato-shi@kirin.co.jp, Tel:81-45-788-7360, Fax:81-45-788-0402)			
FEATURES	Location/Qualifiers			
source	1. 5048 /organism="Schizosaccharomycetes pombe" /mol_type="genomic DNA" /db_xref="taxon:4896" /strain="972h" /db_xref="taxon:4896" /chromosome="II" /map="III" /clone="cosmid c1271" 1307. .3367 /gene="att3+"			
gene	join(1203. .1302,1362. .3511) /gene="att3+" join(1203. .1302,1362. .3511) /gene="att3+" /notes="similar to S. cerevisiae, mouse and human homologue" /codon_start=1 /product="oligosaccharyltransferase subunit" /protein_id="BA076479.1" /db_xref="GI:4568803"			
CDS	/translation="MANSATITTSKKGVSKHQKWKIPLKVLILICIAVASVSRFSV IRYESIHFDPMFPMFRASKILVEGFYNLWDFERSWYPLGRVAGTLYPLGLAVTS GIFKVLHLIRINVINRDCVILLAPAFGITAITYLARELXSDAGCLLAFAFGVIA PGTYSVAGSDVNRATIAITLMSFALMIKAVKSGSPWGACTGLLYFVMTGAWG VFNTNPLHVFLLMGRYTSKLYIVTVYVIGTGLASMQVFPVGFVMTSEHMSA LGVFGGLQFAPFVYKGLVSSKQFQILIRFALVCLGLATVLFALSTGVIAWPTG RFVSLMDTWAKTHPIITASVSEHQPPTWSSLFEDQLQLLPLVGVYLPFKELSEH VFYLIIPVIGTFCGVWRVLUTLTPVCIAAAVAISILLDTWGFVEEDKVSSEA SAKSNKNGISSILSPFTTSGKNIIGISLLSRVLSISFAYLIMFVHSSWVTGNAY SSPTVLTNLNDGSLMYIDDPREAYDLWRNTPYDTKVMWMDYQIAGMADRITL VDNNTWNTTATVNGKMSPEEKAYPLRKHDVDYILIIYGCTLGYSDDMKFLWM IRISQGLWDEIVERNFFTPNGEYTDAAATPWRESLLYKMSYHGAKWLPFNGQYD RANOKLPKSDPOLFTIEAFTTVHVLVLYKKVKKPDTLGRDLKQVTLFEKGRKKS VLOKHLKFL" intron 1303. .1361 /gene="att3+"			
ORIGIN	Alignment Scores: Pred. No.: 1.35e-23 Length: 5048 Score: 32.00 Matches: 32 Percent: 100.00% Conservative: 0 Best Local Similarity: 100.00% Mismatches: 0 Query Match: 4.46% Indels: 0 Dbs: 8 Gaps: 0			
US-10-028-384-6 (1-718) x AB015232 (1-5048)				
QY	322	IlealabrotrphrGlyArgPheTySerLeuTrpAspThrAsnTyAlaLysIleHis	341	
Db	2246	ATCGCTCTGGACAGAGCTTTCATTCTCTTTGGGATACAACTACGCCAAGATTCAT	2305	
QY	342	IleerollerlealaserValserGluHisGlnPro	353	
Db	2306	ATTCTATCATCTGCTCGGTATCAGAACATCAGCCT	2341	
RESULT 14 SPBC1271/c				

```

/ gene="SPBC1271.15c"
1307. 3347
/ gene="SPBC1271.15c"
/ note="translation initiation factor if-2 (predicted);
localization mitochondrion (predicted); non-essential
(PMID 12618370); similar to S. cerevisiae YOL023W"
/ codon_start=1
/ protein_id="CAA22205.1"
/ db_xref="GI:3925784"
/ db_xref="GOA:O59683"
/ db_xref="Swiss-Prot:O59683"
/ translation="MPSLRCTQTHFFHSSPMATKFRFPKSTLTLPVLSVASPA
SLADISVKNHNNHNPOLRQCTHFFHSSPMATKFRFPKSTLTLPVLSVASPA
NLQIPTRLDRLNIGIDVPEYLITYEYSSLAEEYDFDQEAAYGAPSPKSL
NSKSDGSLPVPVLLMGHVDHGKTLTLLDAFKSTIASTHGKITKGAFTPFDKG
SKFTFLDTPCHMAFEAMRRKRGANIADIVLVVAGDDGVKQPTVEAIHQISADVPV
VALTKRDPGPTPIHKIYEOLLNNGIOVEALGGETOIPIISAKTGKIGPELAAILTLA
EIMRIRASPDPPQGVIVSSVTKGSSATVVVKGTVKGMVLVAGKSWCKVRSVL
DVNKSILKQVLPQRAQVYCKDLPAGDLAYEVKESSEAKELSDIYRQNEQNFVE
LAESQNRQVSLAAKSGPAALQETSYSKSFNIAKCEDTGSNEALSDYLKELQFG
KRSRVLYTGVGVPTEDIERASTSDAILISFGVSVKATPRLAEKHNNVLLFHNVI
HLMDDVRKLFALRPPLVQVVTGEAIIISAFIDIKAKRAVHVAGRVNTGTHKSH
IRLVNRDKIITWSGEISLKLKEEVTISIKKRECGILLKNFDEIVTGDKLQTFVEEYK
PPDP"
1811. 2815
/ gene="SPBC1271.15c"
/ note="Pfam match to entry PF00009 GTP_EFTU, Elongation
fac tor Tu family"
1814. 2815
/ gene="SPBC1271.15c"
/ note="Match to GTP_EFTU , Score 114.20"
3215. 3283
/ gene="SPBC1271.15c"
/ note="PS01176 Initiation factor 2 signature"
/ complement(3720.5057)
/ gene="SPBC1271.14"
/ complement(3720.5057)
/ gene="SPBC1271.14"
/ EC number="2.3.1.35"
/ note="amino-acid N-acetyltransferase (predicted);
glutamate N-acetyltransferase (predicted); involved in
ornithine biosynthesis (predicted); involved in arginine
biosynthesis (predicted); amino acid N-acetyltransferase
activity; glutamate N-acetyltransferase activity;
localization mitochondrion matrix (predicted);
non-essential (PMID 12618370); similar to S. cerevisiae
YMR062C
PATH:map00220
aspect:P; term=ornithine biosynthesis; GOid=GO:0006592;
evidence=ISS; db_xref=GOC:unpublished; with=SPTR:Q04728;
date=20030616
aspect:P; term=arginine biosynthesis; GOid=GO:0006526;
evidence=ISS; db_xref=GOC:unpublished; with=SPTR:Q04728;
date=20030616
aspect:P; term=amino acid N-acetyltransferase;
GOid=GO:0004042; evidence=ISS; db_xref=GOC:unpublished;
with=SPTR:Q04728; date=20030616
aspect:P; term=glutamate N-acetyltransferase;
GOid=GO:0004358; evidence=ISS; db_xref=GOC:unpublished;
with=SPTR:Q04728; date=20030616
aspect:C; term=mitochondrial matrix; GOid=GO:0005759;
evidence=ISS; db_xref=GOC:unpublished; with=SPTR:Q04728;
date=20030616
/ codon_start=1
/ protein_id="CAA22204.1"
/ db_xref="GI:3925783"
/ db_xref="GOA:O94345"
/ db_xref="TrEMBL:O94345"
/ translation="MOTKIFARLAGIGRPFVGTMDKTKYFKISSTYFQGSFLNGIA
SGVXANGKQDIALIFSPPCNAAAVFTKNAPQAPVQVRQILNCGGKDIHCVFVNS
GCANAVTGGEGMLAQITAEADNLTPHTWTSNSENSEFPSPSVNVTGIVGQRLKLD
KIQSGLEHAEVDLGSSTHEYWMRAEAICTTDTTFFKLVSRELGIAGKVYRIAGFAGG
MINPNLATLGLFVTDAPISVDVRSILRHAINNSFNSISIDGTSTNTDITAFARKA
AGGSBITKSPAYKEIRDVTDIAQKLAVRDGEGATKFTVQVRGARSEKDAALV
ASTIENSALVKTAFEGEDANWGRILCAVYSGAAVNPATTVSFIPADGTEPLKLVN
GSPQNVDTBASEILSQDELTVDDVLDGGFPYAKNTWTCDFSYDVYRINADYRS"
complement(3723.4946)
/ gene="SPBC1271.14"
/ note="Match to ArgJ , Score 840.54"
/ complement(6265.6931)
/ note="SPMEIRNA D31852 Fission yeast DNA for meirna,
complete sequence"
complement(join(7889.8438,8502.8575))
/ gene="SPBC1271.13"
complement(join(7889.8438,8502.8575))
/ note="SPBC1271.13"
/ note="mitochondrial ribosomal protein L17; essential
(PMID 12618370); similar to S. cerevisiae YUL063C"
/ codon_start=1
/ protein_id="CAA22203.1"
/ db_xref="GI:3925782"
/ db_xref="GOA:O94345"
/ db_xref="TrEMBL:O94345"
/ translation="MKTLYTRKLGRTSAHRQALLNLVTSLVKHESIQTWAKAEA
QRFABKLTMAKRAPNNRKGAEVFEKTLTKVFDVLPVRYNRCGVTLLK
LPRTDNAPGVLEFVDGPKDIEFHTAKTVGICLAQKALPTIRKNIHKVLLFRK
NGKAEFDQIVQKESEHARLKHDEDEKTVKDKMGKGDPIPRPTV"
complement(8156.8437)
/ gene="SPBC1271.13"
/ note="Match to Ribosomal_L17 , Score 114.95"
complement(8439.8454)
/ gene="SPBC1271.13"
/ note="ctaataatgacagacag, splice branch and acceptor"
complement(8496.8501)
/ gene="SPBC1271.13"
/ note="gtagtg, splice donor sequence"
complement(8899.9327)
/ note="RNA from AB016222"
complement(join(9328.9869,9929.10241,10424.10546,
10690.10768,10941.10966,11042.11054,11105.11175))
/ gene="kes1"
/ note="oxysterol binding protein (predicted); involved in
ergosterol biosynthesis (predicted); similar to S.
cerevisiae YHR001W and YOR003W and YOR237W and YPL145C;
similar to S. pombe SPBC646.08c and SPBC354.07c and
SPCC336.01; localization golgi (predicted); non-essential
(PMID 12618370)"
/ codon_start=1
/ protein_id="CAA22202.1"
/ db_xref="GI:3925781"
/ db_xref="GOA:O74178"
/ db_xref="TrEMBL:O74178"
Alignment Scores:
Pred. No.: 8,93e-23 Length: 40907
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.46% Indels: 0
DB: 8 Gaps: 0
US-10-028-384-6 (1-718) x SPBC1271 (1-40907)
Qy 322 IleAlaProThrGlyArgPheTySerLeuTipAspThrAsnTyAlaLysIleHis 341
Db 36751 ATCGCTCCITGGACAGACGTTCTTCTCTTTGGGATACAACTACGCCAGATTTCAT 36692
Qy 342 IleProIleIleIleIleValSerGluHisGlnPro 353
Db 36691 ATTCCATCATCTTCCTCGGTATCAGAACATCAGCCT 36656

```

RESULT 15

CR382134_03

WPCOMMENT

Sequence split into 14 fragments LOCUS CR382134 Accession CR382134

Fragment Name	Begin	End
CR382134_00	1	110000
CR382134_01	100001	210000
CR382134_02	200001	310000
CR382134_03	300001	410000
CR382134_04	400001	510000
CR382134_05	500001	610000
CR382134_06	600001	710000
CR382134_07	700001	810000
CR382134_08	800001	910000
CR382134_09	900001	1010000
CR382134_10	1000001	1110000
CR382134_11	1100001	1210000
CR382134_12	1200001	1310000
CR382134_13	1300001	1349926

Continuation (4 of 14) of CR382134 from base 300001 (CR382134 Debaryomyces hansenii chrC

Alignment Scores:

Pred. No.:	2,18e-22	Length:	110000
Score:	32.00	Matches:	32
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	4.46%	Indels:	0
DB:	8	Gaps:	0

US-10-028-384-6 (1-718) x CR382134_03 (1-110000)

Qy	322	IleAlaProThrThrGlyArgPheTyrsSerLeuTyrAspThrAsnTyAlaLysIleHis	341
Db	65525	ATTGCACCTTGGACTGGCGTTCTATTCTTTATGGGATCTAACTATGCCAAGATTTCAC	65584
Qy	342	IleProIleIleAlaSerValSerGluHisGlnPro	353
Db	65585	ATTCTATTAATTCGATCCGTTTCTGACATCAACCT	65620

Search completed: December 14, 2004, 22:41:46

Job time : 8286.53 secs

Handwritten signature or initials, possibly reading "KJ" or "KJL".

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 827.256 Seconds
(without alignments)
4556.130 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 718

Sequence: 1 WSDRSCVLSVFQIKLVI.....TRSTKTRRSIKRPELGLRV 718

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO spool/US10028384/runat 14122004 131645 6790/app_query.fasta.1.3740
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=Oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn_1.1986@runat.14122004 131645 6790 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -XGAPOP=6
-XGAPEXT=7 -XGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseq 23Sep04:*

1: geneseqn1980s:.*
2: geneseqn1990s:.*
3: geneseqn2000s:.*
4: geneseqn2001as:.*
5: geneseqn2001bs:.*
6: geneseqn2002as:.*
7: geneseqn2002bs:.*
8: geneseqn2003as:.*
9: geneseqn2003bs:.*
10: geneseqn2003cs:.*
11: geneseqn2003ds:.*
12: geneseqn2004as:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	718	100.0	2733	10	ADD94787 Yeast STT
2	605	84.3	2000	10	Acc61293 Gene sequ
3	605	84.3	2000	10	Adk63659 Disease t
4	121	16.9	381	6	Abn23953 Human ORF
5	32	4.5	2256	6	Abz23036 Candida a
6	21	2.9	1848	8	Abt18969 Aspergill

7	21	2.9	1969	8	ABT18375
8	21	2.9	2232	8	ABT20789
9	21	2.9	2466	10	Adb69300 C. neofo
10	21	2.9	2603	8	ABT20131
11	21	2.9	2882	10	ADb69539
12	21	2.9	3969	8	ABT17781
13	21	2.9	4603	8	ABT19595
14	21	2.9	4738	10	ADb69178
15	20	2.8	2417	10	ADD94789
16	20	2.8	2699	4	ABL13247
17	20	2.8	2953	4	ABL18224
18	20	2.8	4922	4	ABL13246
19	17	2.4	313	8	ABZ20328
20	17	2.4	387	4	ABA08357
21	17	2.4	393	4	ABA08547
22	17	2.4	558	5	AAF93612
23	17	2.4	1209	4	AAH33264
24	17	2.4	1209	6	ABL89850
25	17	2.4	1543	2	AAH85055
26	17	2.4	1543	8	ACD18981
27	17	2.4	1543	12	ADG78372
28	17	2.4	1543	12	ADG78372
29	17	2.4	1664	4	AAK94184
30	17	2.4	1664	12	ADL30661
31	17	2.4	2481	10	ADD94783
32	17	2.4	2710	10	ADD94785
33	17	2.4	2855	4	ABL02795
34	17	2.4	6153	4	ABL02794
35	16	2.2	154	3	AAC93955
36	16	2.2	245	3	AAC17015
37	16	2.2	247	2	AAH40252
38	16	2.2	474	5	AAF93633
39	16	2.2	485	9	ACH24071
40	16	2.2	500	6	ABV88332
41	16	2.2	1114	4	AAH99794
42	16	2.2	1371	12	ADP28508
43	16	2.2	2472	10	ADD94793
44	16	2.2	2472	10	ADH28844
45	16	2.2	2760	5	ABV24502

ALIGNMENTS

RESULT 1
ADD94787

ID ADD94787 standard; DNA; 2733 BP.

XX AC ADD94787;

DT 29-JAN-2004 (first entry)

DE Yeast STT3 gene sequence.

XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide; yeast;
KW gene; ds; STT3.

XX OS Saccharomyces cerevisiae.

XX PN WO2003054008-A2.

XX PD 03-JUL-2003.

XX PF 18-DEC-2002; 2002WO-CA001967.

XX PR 20-DEC-2001; 2001US-00028384.

XX PA (COMP-) COMPATIGENE INC.

XX PI Perreault C, McBride K;
 XX DR WPI; 2003-559122/52.
 XX DR P-PSDB; ADD94788.
 XX PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 XX PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 XX PT or breast cancer, or for suppressing an immune response in an autoimmune
 XX PT disease.
 XX PS Disclosure; SEQ ID NO 5; 66pp; English.
 XX CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the yeast S133 gene which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX SQ Sequence 2733 BP; 712 A; 597 C; 583 G; 841 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 2733
 Score: 718.00 Matches: 718
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0
 US-10-028-384-6 (1-718) x ADD94787 (1-2733)
 QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleuLeuValle 20
 DB 219 ATGGGATCCGACCGGTCGTGTGTTTCTGTCTTTTCAGACCATTCCTCAAGCTCGTCATC 278
 QY 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 DB 279 TTCGTGGCGATTTTGGCGTGCATATCATCATCAGTTTGTTCAGTCATCAAAATTTGAG 338
 QY 41 SerIleIleHisGluPheAspProTTPheAsnTyrArgAlaThrIleLysPheValasn 60
 DB 339 TCTATATCCAGTAATTCGACCCCTGGTTCAATATAGGGCTACCAAAATATCTCTCAAC 398
 QY 61 AsnSerPheTyrLysPheLeuAsnTTPheAspAspArgThrTTPyrProLeuGlyArg 80
 DB 399 AATTCGTTTACAAGTTTGAACCTGGTTTGAACCGCTACCTGGTACCCCTCGGAAGG 458
 QY 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrPheIle 100
 DB 459 GTTACTGGAGGAGCTTATATATCTCTGGTTTGTATGACACTAGTGCCTTCATCTGGCAGCC 518
 QY 101 LeuArgAsnTTPLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlapro 120
 DB 519 CTGCGCAACTGTTGGGCTTGCCTTGCATATGACATCAGAAACGTTGTGTGCTATTTTCGCCA 578
 QY 121 LeuPheSerGlyValThrAlaThrAlaThrTyrGluPheThrIleLysAspAla 140
 DB 579 CTATTTTCTGGGTCACCGCTGGCGACTTACGATTTACGAGAGATTAAGATGCC 638
 QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160

DB 639 AGCGTGGGCTTTTGGCTGCTGGTTTATAGCCANTGTCCCGTTATATATATAGATCA 698
 QY 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 DB 699 GTGGCGGGCTCTACGATAATAGAGCCATTCGCACTATTAATATGCTCACTTTTCATG 758
 QY 181 PheTTPleuValAlaGlnThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 DB 759 TTTTGGATTAAGCCCAAAAGACTGGCTCTATCATGACGCAACGTTGTCAGCTTTATTC 818
 QY 201 TyrPheTyrMetValSerAlaTTPGlyGlyTyrValPheIleThrAsnLeuLeuProLeu 220
 DB 819 TACTTCTACATGCTGGCTTGGGCTGGATAGTGTTCATCACCACATTCATCCACATC 878
 QY 221 HisValPheLeuLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 DB 879 CATGCTTTTGTGATTTTGTATGGGAGATATTCGTCACAACTGATATTCGCTACACC 938
 QY 241 ThrTTPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 DB 939 ACTTGGTACGCTATTTGGAACCTTTCATCCATGAGATCCCATTTGCTGCTTTCTACCT 998
 QY 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 DB 999 ATCAGGTCTACGACCATTCGCGCATTTGGGTGTTTTCGGTTTGTATTCAGATTTGCGC 1058
 QY 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 DB 1059 TTCGTGTGATCTCGTGAAGGGCCAAATCAGCACAGCTAAAGTTTAAAGTCATCATGATG 1118
 QY 301 SerLeuPheLeuLeuLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 DB 1119 TCTCTGTTTGTGATCTTGGTCTTGGTGTGTGGCTTCTGCGCTTTCGCTATATATGCGG 1178
 QY 321 LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTTPAspThrAsnTyrAlaLysIle 340
 DB 1179 TTGATTTGCCCTTGGCTAGTATTTATTCGTTATGGGATACCAACTACCGAAAGATC 1238
 QY 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTTProAlaPhePhe 360
 DB 1239 CACATTTCTATCATTTGCTCTCGTTTCGAAACATCAACCCGTTTCGTGGCGCTTCTTTC 1298
 QY 361 PheAspThrHisPheLeuIleTTPLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 DB 1299 TTTGATACCCACTTTTGTATCTGCTATTTCCCGCGGTGATTCCTACTATTTCTTCGAC 1358
 QY 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuLysSerTyrPheAlaGly 400
 DB 1359 TTGAAAGACGACGACGTTTTTGTCTATCGTTTCTCGTTCTGTGTTCTGCTATTTGCGG 1418
 QY 401 ValMetValArgLeuMetLeuLeuLeuThrProValIleCysValSerAlaAlaValAla 420
 DB 1419 GTTATGGTTAGATTGATGTTGACTTTGACACCATCATCTGTGTGTCGCCGCGCTCGCA 1478
 QY 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
 DB 1479 TTGTCCAAAGATATTGACATCTACCTGATTTCAAGACAAAGTACCGCAAAATACGCCATC 1538
 QY 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
 DB 1539 AAACCTGGGACACTACTGGCCAAATTTGTTTTCGGATCATCTCTTTATTTTGTAT 1598
 QY 461 LeuPheValPheHisSerThrTTPyrValThrArgThrAlaTyrSerSerProSerValVal 480
 DB 1599 CTTTTCTGCTTCCATCTCTACTTGGGTAAACAAAGACTGACATCTCTCTCTCTCTCTCT 1658
 QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
 DB 1659 TTGCCATCACAAACCCAGATGTTAATTTGGCTGTGATCGACGACTTCAGGGAAGCGTAC 1718
 QY 501 TyrTTPLeuArgMetAsnSerAspGluAspSerLysValAlaIleAlaIleTTPTrpAspTyrGly 520

Db 1719 TATTGTTAAGATGAACCTCTGATGAGCAGACAGTAAGTTGTCAGCGTGGTGGATTACGGT 1778
 Qy 521 TyrGlnIleGlyClyMetAlaAspArgThrThrLeuValAspAsnAenThrTrpAsnAen 540
 Db 1779 TACCAAAATGGTGGCAGTGCAGACAGAACCACTTTAGTCGATACAAACACGTCGGAACAAT 1838
 Qy 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTrpGluIle 560
 Db 1839 ACTCACATCGCCATCGTTGTTAAAGCCATGGCTTCCCTGAAGAGAATCTTACGAAAT 1898
 Qy 561 LeuIleGluHisAspValAspTrpValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
 Db 1899 CTAAAGACATGATGTCATATGTCGTGTCATCTTGTGTGTCTAAATGGGTGGT 1958
 Qy 581 GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu 600
 Db 1959 GGTGATGATCAACAAATCTTGTGGATGATCAGATTAGCGAGGAACTCTGCCAGAA 2018
 Qy 601 GluIleIleGluArgTrpPheTrpThrAlaGluGlyGluTrpArgValAspAlaArgAla 620
 Db 2019 GAGTAAAGAGCGTTATTTCTATCCGAGAGGGAGATACAGAGTAGATGCAAGGCT 2078
 Qy 621 SerGluThrMetArgAsnSerLeuLeuTrpLysMetSerTrpLysAspPheProGlnLeu 640
 Db 2079 TCTGAGACCATGAGGAACCTCGTACTTTTACAAGATGTCCTACAAAGATTTCCACAAATTA 2138
 Qy 641 PheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAspVal 660
 Db 2139 TTCATGGTGGCCAGCCACTGACAGAGTGCGTCAACAAATGATCACCATTAGACGTC 2198
 Qy 661 ProProLeuAspTrpPheAspGluValPheThrSerGluAsnTrpMetVala:gilLeTyr 680
 Db 2199 CCACCATTAGACTACTTCGACGAAGTTTCTTACTTCCGAAAACCTGGATGGTTAGAAATAT 2258
 Qy 681 GlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
 Db 2259 CAATTGAAGAAGATGATGCCACAGTAGAATCTTGAGGACGTTGGTGGTTAACCAAG 2318
 Qy 701 SerSerThrLysThrArgSerIleLysArgProGluLeuGlyLeuArgVal 718
 Db 2319 TCTTCTACAAACACAGAGGTCCATAAAGACCTGAATTAGGCTTTGAGAGTC 2372
 RESULT 2
 ACC61293
 ID ACC61293 standard; DNA; 2000 BP.
 XX
 AC ACC61293;
 XX
 DT 20-JUN-2003 (first entry)
 XX
 DE Gene sequence #SEQ ID 1368.
 XX
 XX Multiprotein complex; eukaryote; drug target; diagnosis; gene; ds.
 XX
 OS Saccharomyces cerevisiae.
 XX
 PN EP1258494-A1.
 XX
 PD 20-NOV-2002.
 XX
 XX 20-DEC-2001; 2001EP-00130253.
 XX
 XX 15-MAY-2001; 2001EP-00111774.
 XX
 PA (CELL-) CELLZONE AG.
 XX
 XX Bauer, A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
 PI Marzioch M, Schultz JD, Superti-Furga GD;
 XX
 DR WPI; 2003-250078/25.
 DR P-PSDB; ABR53251.
 XX
 PT New isolated protein complexes useful for diagnosing a disease or

PT disorder, or as a target for an active agent of a pharmaceutical,
 PT preferably a drug target in the treatment or prevention of disease or
 disorder.
 XX
 XX Disclosure: SEQ ID NO 1368; 17pp + Sequence Listing; English.
 XX
 CC The invention relates to multiprotein complexes from eukaryotes. Proteins
 CC of the invention and DNA sequences encoding them are given in records
 CC AB952568-AB953903 and ACC610-ACC61944 respectively. The complexes are
 CC obtainable by using a protein as a bait and isolating the set of proteins
 CC which is attached thereto from cells. Such protein complexes may comprise
 CC up to 30 distinct proteins. Protein complexes of the invention are useful
 CC for diagnosing a disease or disorder, or as a target for an active agent
 CC of a pharmaceutical, preferably a drug target in the treatment or
 CC prevention of a disease or disorder. Note: The sequence data for this
 CC patent is not represented in the printed specification, but is based on
 CC sequence information supplied by the European Patent Office. The complete
 CC document is available on CD-ROM
 XX
 SQ Sequence 2000 BP; 461 A; 459 C; 453 G; 627 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 2000
 Score: 605.00 Matches: 665
 Percent Similarity: 99.70% Conservative: 0
 Best Local Similarity: 99.70% Mismatches: 1
 Query Match: 84.26% Indels: 2
 Ds: 10 Gaps: 0
 US-10-028-384-6 (1-718) x ACC61293 (1-2000)
 Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 Db 1 ATGGGATCCGACCGCTGCTGTTTGTCTGTGTTTCAGACCATCTCTCAAGCTGTCATC 60
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 61 TTCGTGGCGATTTTGGGGCTGCCATCATCATCACGTTGTTGTCAGTCAATCAATTTGAG 120
 Qy 41 SerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrLysTrpLeuValAsn 60
 Db 121 TCTATTATCCATGAATTCGACCCCTGGTTCAATTTAGGGCTACCAATATCTCTCAAC 180
 Qy 61 AsnSerPheTrpLysPheLeuAsnTrpPheAspAspArgThrTrpTrpProLeuGlyArg 80
 Db 181 AATTCGTTTACAAAGTTTGTGAACCTGGTTGACGACCGTACCTGTACCCCTCGGAAGG 240
 Qy 81 ValThrGlyGlyThrLeuTrpProGlyLeuMetThrThrSerAlaPheIleTrpHisAla 100
 Db 241 GTTACTGGAGGACTTTATATCTCTGTTTGTATGACGACTAGTGCCTTCATCTGGCAGCC 300
 Qy 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 Db 301 CTGGCAACTGGTTGGCTTGGCCATTCATGACATCAGAAACGTTGTGTCTATTTCGCCCA 360
 Qy 121 LeuPheSerGlyValThrAlaTrpAlaThrTrpGluPheThrLysGluIleLysAspAla 140
 Db 361 CTATTTCTGGGTCACCGCTGGGCGACTTACGAATTTACGAAAGAGATTAAAGATGCC 420
 Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTrpLysSerArgSer 160
 Db 421 AGCGCTGGGCTTTGGCTGCTGTTTATAGCATTTGCCATTTGCCGTTATATCTAGATCA 480
 Qy 161 ValAlaGlySerTrpAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 481 GTGGCGGGCTCTACGATAATGAGGCCATTGGCATTACACTATTAAATGTCACATTTCATG 540
 Qy 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 Db 541 TTTTGGATTAAAGCCCAAGAGACTGCTCTATCATGACGACCAACGCTGTGACGCTTATTC 600
 Qy 201 TyrPheTrpMetValSerAlaTrpGlyGlyTrpValPheIleThrAsnLeuIleProLeu 220

Db 601 TACTTCTACATGGTGTGGGTGGGATGATCGTGTTCATCACCACCTTGATCCCACTC 660
 QY 221 HisValPheLeuLeuLeuLeuMetGlyArgTyrSerSerIysLeuTyrSerAlaTyrThr 240
 Db 661 CATTGCTTTTGGTGAATTTGATGGGAGATATTCGTCCAACTGATTTCTCCCTACACC 720
 QY 241 ThrTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 721 ACTTGGTACCGCTATTGGAACCTGTTCATCCATGCAGATCCCATTTGCGGTTCTACCT 780
 QY 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 781 ATCAGGCTTACAGACCATGGCCGCAATGGGTTGTTTCGTTTGAATCGAATGTCGCC 840
 QY 281 PheGlyAspPheValIysGlyGlnIleSerThrAlaIysPheIysValIleMetMetVal 300
 Db 841 TTCCGTGATCTTCGTGAAGGGCCAAATCAGCACAGCTAAAGTTAAAGTCATCATGATGTT 900
 QY 301 SerLeuPheLeuLeuLeuValLeuGlyValIleGlyLeuSerAlaLeuThrTyrMetGly 320
 Db 901 TCTCTGTTTGTATCTGGTCCCTTGGTGGTGGTCTTCTGCTTGCCTATATGGGG 960
 QY 321 LeuIleAlaProTyrThrGlyArgPheTyrSerLeuTyrAspThrAsnTyrAlaIysIle 340
 Db 961 TTGATTTGCCCTTGGACTGTGGAGATTTTATTCGTTATGGGATACCAACTACGCAAGATC 1020
 QY 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTyrProAlaPhePhe 360
 Db 1021 CACATTCCTATCATGTGCTCGTTCGCAACATCAACCCGTTTCGTGGCCGCTTCTTC 1080
 QY 361 PheAspThrHisPheLeuIleTyrLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 Db 1081 TTGTATACCCATTTTGAATCGCTATTCGCCGCGGTGATTCCTACTATTCCTCGAC 1140
 QY 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
 Db 1141 TTGAAGACGACGACGCTTTTGTTCATCGCTTCTCGCTTCTGTTCTGTTCTCGGT 1200
 QY 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaIleValAla 420
 Db 1201 GTATGGTTAGATGTGATGTTGACTTTGACACCAAGTATGTTGTGTCGCCGCGCTCGCA 1260
 QY 421 LeuSerLysIlePheAspIleTyrLeuAspPheIysThrSerAspArgLysTyrAlaIle 440
 Db 1261 TTGTCCAAAGATTTGACATCTACCTCGAATTCAGACAGATGACCGCAATACGCCATC 1320
 QY 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
 Db 1321 AAACCTGCGCACTACTGCGCAATGATTTGCTCGGATCATCTCTTTTATTGTGAT 1380
 QY 461 LeuPheValPheHisSerThrTyrValThrArgThrAlaTyrSerSerProSerValVal 480
 Db 1381 CTTTTCGCTTCCATCTCTACTTGGTAACAAGAACTGCATCTCTCTCTCTCTCTGTT 1440
 QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
 Db 1441 TTGCCATCACAACCCCAAGATGGTAATTTGGGTTGATGACGACTTCAGGAACGATAC 1500
 QY 501 TyrTrpLeuArgMetAsnSerAspGluAspSerIysValAlaAlaTyrTrpAspTyrGly 520
 Db 1501 TATTGGTTAAGAAATGAATCTGTATGAGCAGATGAAGTTGACGGTGGTGGGATACCGT 1560
 QY 521 TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
 Db 1561 TACCAAAATGGTGGCATGGCAGACAGAACCATTTAGTCGATACACACGTTGGAACAT 1620
 QY 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluIleLysSerTyrGluIle 560
 Db 1621 ACTCATCTGGCATCTGTTGGTAAAGCCATGCTTCCCTGAAAGAAATCTTACCAAAAT 1680
 QY 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
 Db 1681 CTAAAGACGATGATGCTATGTTCTTGGTCTCTTGGTCTCTTGGTCTTGGTCTTGGT 1740

QY 581 GlyAspAspIleAsnLysPheLeuTyrMetIleArgIleSerGluGlyIleTyrProGlu 600
 Db 1741 GGTGATGATCATCAACAATCTTGTGATGATCAGATAGCGGGGATCTGGCCAGAA 1800
 QY 601 GluIleLysGluArgTyr-PheTyrThrAlaGluGlyGluTyrArgValAlaArgAl 620
 Db 1801 GAGATAAAGAGCGTGA-TTTCATATACCGCAGAGGAGATACAGAGTAGATGCAAGGCG 1859
 QY 620 aserGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnIle 640
 Db 1860 TTCTGAGCCATGAGGAACCTGCTACTTTACAGATGTCTCTACAAAGATTTCCCAATT 1919
 QY 640 uPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa 660
 Db 1920 ATTCAATGTCGCCAAGCCACTGACAGTGCCTCAACAAATGATCACACCATTAGACGT 1979
 QY 660 lProProLeuAspTyrPhe 666
 Db 1980 CCCACCATTAGACTTCTC 1998

RESULT 3

ADK63659 standard; DNA; 2000 BP.

ID ADK63659

AC ADK63659;

XX 06-MAY-2004 (first entry)

DT Disease treating protein complex-derived gene #824.

XX protein complex; drug target; diagnosis; gene; ds.

DE Unidentified.

XX EP1338608-A2.

XX 27-AUG-2003.

XX 20-DEC-2002; 2002EP-00102902.

XX 20-DEC-2001; 2001EP-00130253.

XX (CELL-) CELLZOME AG.

PA Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;

PI Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;

PI Michon A, Leutwein C, Rick J;

XX MPI; 2003-638460/61.

DR P-PSDB; ADK63658.

XX New proteins and protein complexes from eukaryotes, useful as targets in

PT drug screening, or in diagnosing or screening for the presence of a

PT disease or disorder, or a predisposition for developing a disease or

PT disorder in a subject.

XX Disclosure; SEQ ID NO 1648; 13pp; English.

PS The invention relates to novel protein complexes comprising a first and a

CC second protein, or its derivative, fragment, homologue or variant. The

CC proteins are selected from given protein complexes, which are not defined

CC in the specification. The variants are encoded by nucleic acids that

CC hybridize to the nucleic acids encoding the proteins under low stringency

CC conditions. The protein complexes are useful as targets for an active

CC agent of a pharmaceutical. These protein complexes are particularly

CC useful as drugs targets for the treatment or preventing of a disease or

CC disorder. The complexes and methods above are useful in diagnosing or

CC screening for the presence of a disease or disorder or a predisposition

CC for developing a disease or disorder in a subject. These are also useful

CC in screening for a drug for treatment or prevention of a disease or

CC disorder. The molecule that modulates the amount, activity or protein

CC components of the complex is useful for the manufacture of a medicament

CC for the treatment or prevention of a disease or disorder. This sequence
 CC corresponds to a gene of the invention. (Note: the sequence data for this
 CC patent did not form part of the printed specification but was obtained
 CC from the EPO in electronic format).

XX
 SQ Sequence 2000 BP; 461 A; 459 C; 453 G; 627 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2000
 Score: 605.00 Matches: 665
 Percent Similarity: 99.70% Conservativity: 1
 Best Local Similarity: 99.70% Mismatches: 0
 Query Match: 84.26% Indels: 2
 DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADK63659 (1-2000)

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 DB 1 ATGGGATCCGACCGGTCGTGTTTGTCTGTGTTTCAGACCATCTCCAAAGCTCGTCATC 60
 QY 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 DB 61 TTCGGGGCGATTTTGGGGCTGCCATATCATCAGCTTGTGTTGAGTCATCAAAATTGAG 120
 QY 41 SerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrLysTyrLeuValAsn 60
 DB 121 TCTATTATCCATGAATTCGACCCCTGGTTCAATTATAGGCTACCAAAATATCTCGTCAAC 180
 QY 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArg 80
 DB 181 AATTGGTTTTCAGAGTTTTCAGACCTGGTTTTCAGACCTGGTACCTGGTACCTGGTACCT 240
 QY 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla 100
 DB 241 GTTACTGGAGGACATTATATCTGTTTGTATGACGACTAGTGGTTCATCTGGCAGCC 300
 QY 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 DB 301 CTGGCACTGGTGGGCTTGGCCATTGACATCAGAACGTTTGTGCTGTTATTTGGCCCA 360
 QY 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 DB 361 CTATTCTTGGGGTCCACCGCTGGGCGACTTACGAATTTACGAAAGAGATTAAGATGCC 420
 QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 DB 421 ACGGCTGGGCTTTTGGCTGGTGTATAGCCATTGTCCTGGTTATATCTAGATCA 480
 QY 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 DB 481 GTGGGGGCTCTACGATAATGAGGCCATTGCCATTACATATTATATGGTCACTTTTCATG 540
 QY 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 DB 541 TTTTGGATTAGGCCCAAAAGACTGGCTCTATCATGACGCAACGTCGTCAGCTTTATTC 600
 QY 201 TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
 DB 601 TACTTCTACATGGTGTGGCTGGGTGGATACGGTTCATCAACCACTTGTATCCCACTC 660
 QY 221 HisValPheLeuLeuIleLeuMetClyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 DB 661 CATGCTTTTGTCTGATTGATGGCGAGATATTCTGTCGCAAACTGATTTCTGCTACACC 720
 QY 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 DB 721 ACTTGGTACGCTATTATGGAACGTTTGCATCCATGACAGATCCCATTTGTCGGTTCCTACCT 780
 QY 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 DB 781 ATCAGGTCTACGACCATGCGCGCATTTGGGTGTTTTCGGTTTGTATTCAGATTGTCGCC 840

QY 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetVal 300
 DB 841 TTCGGTACTTCGTGAAGGCCAAATCAGCACAGCTAAGTTTAAAGTCATCATGATGTT 900
 QY 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaIleThrTyrMetGly 320
 DB 901 TCTCTGTTTTTGATCTTGGTCCCTTGGTGGTGGGACTTTCCTGCTTGACCTATATGGGG 960
 QY 321 LeuIleAlaProTrpTrpThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle 340
 DB 961 TTGATTGGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCAACATACGCAAGATC 1020
 QY 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProAlaPhePhe 360
 DB 1021 CACATTCCTATCATTCCTCCGTTTCGAAACATCAACCCGTTTCGGGGCCGCTTCTTC 1080
 QY 361 PheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 DB 1081 TTTGATACCCACTTTTGTGATCTGGCTATTCGCGCGGTGATATTCCTACTATTCTCTCGAC 1140
 QY 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
 DB 1141 TTGAAAGACGAGACGCTTTTGTATCGCTTACTCCGTTCTGTTCTGTTCTGTTCTGCT 1200
 QY 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
 DB 1201 GTTATGGTTAGATTGATTGATCTTTTGACACACAGTCATCTGTGTGTCGCGCGCTCGCA 1260
 QY 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
 DB 1261 TTGTCCAAAGATTTTGTGATCTTACCTGGATTCCAAAGCAAGTGCAGCAAAATACGCCATC 1320
 QY 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
 DB 1321 AAACCTCGGCACCTACTTGGCCAAATGATTGTTTCCGGATCATCTCATCTTTATTTGAT 1380
 QY 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerValVal 480
 DB 1381 CTTTTCGTTCTCCATCTTCTGTTGGTAAAGAACTGCACTCTCTCTCTCTCTCTCTCTCT 1440
 QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
 DB 1441 TTGCCATCACAAACCCAGATGGTAAATGGCTTGCATCGACGCTTACGGAAGCGTAC 1500
 QY 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTyrGly 520
 DB 1501 TATTGGTTAAGATGAACCTCTGATGAGGACAGTAAGGTTGCGGCTGGGATACCGT 1560
 QY 521 TyrGlnIleGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
 DB 1561 TACCAATTTGGTGGCATGGCAGACAGAACCATTTAGTCGATTAACACACGTTGGAACAAT 1620
 QY 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluLysSerTyrGluIle 560
 DB 1621 ACTCACATCGCCATCGTTGGTAAAGCCATGGCTTCCCTCGAAGAGAAATCTTACGAAT 1680
 QY 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyLeuIleGlyPheGly 580
 DB 1681 CTAAAGAGCATGATGCTGATTATGCTTGGTCACTCTTGGTGTCTAATTTGGGTTGGT 1740
 QY 581 GlyAspAspIleLeuLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu 600
 DB 1741 GGTGATGACATCAACAAATTTCTTGTGATGATCGAATTTAGCGGGGAATCTGGCCAGAA 1800
 QY 601 GluIleLysGluArgTyr-PheTyrThrAlaGluGlyGlyTyrArgValAspAlaArgAl 620
 DB 1801 GAGATAAAGAGCGTGA-TTTCTATACCGCAGAGGGAGAAATACAGAGTAGATGCAAGGSC 1859
 QY 620 aSerGluThrMetArgHisSerLeuTyrLysMetSerTyrLysAspPheProGlnLe 640
 DB 1860 TTCTGAGACCATGAGGAACCTGCTACTTTACAAGATGCTCTACAAAGATTTCCCAAT 1919
 QY 640 uPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa 660

DB 1920 ATTCAATGGTGGCCAGCCACTACAGAGTGGCTACAAATGATCACACCATAGAGCT 1979
 QY 660 lProProLeuAspTyrPhe 666
 DB 1980 CCCACCATAGACTACTC 1998
 RESULT 4
 ABN23953
 ID ABN23953 standard; cDNA; 381 BP.
 XX
 AC ABN23953;
 XX
 DT 24-JUN-2002 (first entry)
 XX
 DE Human ORFX polynucleotide sequence SEQ ID NO:16383.
 KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200192523-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US010836.
 XX
 PR 30-MAY-2000; 2000US-0206132P.
 PR 29-AUG-2000; 2000US-0228716P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach MD;
 XX
 DR WPI; 2002-106308/14.
 DR P-PSDB; ABP08201.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders.
 XX
 PS Disclosure; SEQ ID NO 16383; 1037pp; English.
 XX
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-1491 (see table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage. N.B. The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic

CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 381 BP; 100 A; 78 C; 86 G; 116 T; 0 U; 1 Other;
 Alignment Scores:
 Pred. No.: 7.03e-115 Length: 381
 Score: 121.00 Matches: 121
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 16.85% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-6 (1-718) x ABN23953 (1-381)
 QY 452 SerGlySerPhePheLeuTyrLeuPheValPheHisSerThrTrpValThrArg 471
 DB 1 TCCGGATCATTCATCTTTTATTGTTATCTTTTCTTCCATCTCTTGGGTAAACAAGA 60
 QY 472 ThrAlaTyrSerSerProSerValValLeuProSerClnThrProAspGlyLysLeuAla 491
 DB 61 ACTGCATATCTCTTCTCTCTCTGTTGTTGTCATCAAAACCCAGATGTAATTTGGCG 120
 QY 492 LeuIleAspAspPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSer 511
 DB 121 TTGATCGACGACTTCAGGGAAGCGTACTATTGTTAAGAATGAACCTCTGATGAGGACAGT 180
 QY 512 LysValAlaAlaTyrTrpAspTyrGlyClnIleGlyGlyMetAlaAspArgThrThr 531
 DB 181 AAGTTGACAGCGTGGTGGGATTACGGTTACCAAAATGTTGGCATGGCAGACAAACCACT 240
 QY 532 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAla 551
 DB 241 TTAGTCGATAACACACACGTCGAACAATACTCACAATCCATCGTCTGTAAGCCATGGCT 300
 QY 552 SerProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuVal 571
 DB 301 TCCCTGGAAGAGAAATCTTACGAATTTCTAAAGAGCATGATGTCGATTATGTTGGTTC 360
 QY 572 Ile 572
 DB 361 ATC 363
 RESULT 5
 ABN232036
 ID ABN232036 standard; DNA; 2256 BP.
 XX
 AC ABN232036;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE Candida albicans essential gene SEQ ID NO 6323.
 XX
 KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KW signal transduction; DNA replication; cell division; growth;
 KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
 XX
 OS Candida albicans.
 XX
 PN WO200253728-A2.
 XX
 PD 11-JUL-2002.
 XX
 PF 26-DEC-2001; 2001WO-US049486.
 XX
 PR 29-DEC-2000; 2000US-0259128P.
 PR 20-FEB-2001; 2001US-00792024.
 PR 22-AUG-2001; 2001US-0314050P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
 XX
 DR WPI; 2002-566694/50.

DR P-PSDB; ABP73486.
 XX Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele of
 PT a gene and placing other allele of the gene under conditional expression.
 XX Claim 37; SEQ ID NO 6323; 167bp + Sequence Listing; English.
 XX The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter-replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office
 XX
 SQ Sequence 2256 BP; 648 A; 336 C; 456 G; 816 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,658-22 Length: 2256
 Score: 32.00 Matches: 32
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.46% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-6 (1-718) x AB232036 (1-2256)

QY 322 IleAlaProThrThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIleHis 341
 Db 991 ATTGCTCCTTGGACAGGTAGATTTTATTCCTATGGGATACAAATTATGCCAGATTTCAT 1050
 QY 342 IleProIleIleAlaSerValSerGluHisGlnPro 353
 Db 1051 ATTCCAATTATTGCTTCGTCTTCTGACATCAACCT 1086

RESULT 6
 ABT18969
 ID ABT18969 standard; DNA; 1848 BP.

XX AC ABT18969;

XX DT 16-APR-2003 (first entry)

XX DE Aspergillus fumigatus essential gene #1327.

XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 XX cancer; contamination; biofilm; antibody; immune response; ds.
 XX OS Aspergillus fumigatus.

XX PN WC200286090-A2.

XX PD 31-OCT-2002.

XX XX

PF 23-APR-2002; 2002WO-US013142.
 XX
 PR 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287086P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 XX
 PT New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX
 PS Disclosure; Page: 175pp; English.

XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention

SQ Sequence 1848 BP; 363 A; 510 C; 447 G; 528 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4,91e-11 Length: 1848
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x ABT18969 (1-1848)

QY 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 Db 448 TACATGGTGTGGCATGGGTGGGTATGTTCTTATTACGAACCTGATCCCTGCACGTT 507

QY 223 Phe 223

Db 508 TTT 510

RESULT 7

ABT18375

ID ABT18375 standard; DNA; 1969 BP.

XX AC ABT18375;

XX XX

DT XX 16-APR-2003 (first entry)
 DE XX Aspergillus fumigatus essential gene #733.
 KW XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 XX XX cancer; contamination; biofilm; antibody; immune response; ds.
 OS XX Aspergillus fumigatus.
 XX XX WO200286090-A2.
 PN XX 31-OCT-2002.
 PD XX
 XX XX 23-APR-2002; 2002WO-US013142.
 PF XX
 XX XX 23-APR-2001; 2001US-0285697P.
 PR XX 27-APR-2001; 2001US-0287066P.
 PR XX 05-JUN-2001; 2001US-0295890P.
 PR XX 09-JUL-2001; 2001US-0303899P.
 PR XX 31-AUG-2001; 2001US-0316362P.
 XX XX (ELIT-) ELITRA PHARM INC.
 PA XX
 XX XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 PI XX WPI; 2003-093124/08.
 DR XX
 XX XX New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX XX
 PS Disclosure; Page; 175pp; English.
 XX XX
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX
 SQ Sequence 1969 BP; 389 A; 540 C; 480 G; 560 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 5.22e-11 Length: 1969
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0
 US-10-028-384-6 (1-718) x ABT18375 (1-1969)
 QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuLeuProLeuHisVal 222

Db 514 TACATGGTGTGGCATGGGTGGGTATGCTTCTTACGAACCTGATCCCTGCACGTT 573
 QY 223 Phe 223
 Db 574 TTT 576
 RESULT 8
 ABT20789
 ID ABT20789 standard; DNA; 2232 BP.
 XX XX
 AC ABT20789;
 XX XX
 DT 16-APR-2003 (first entry)
 DE XX Aspergillus fumigatus essential gene #3147.
 XX XX
 KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 XX XX cancer; contamination; biofilm; antibody; immune response; ds.
 OS XX Aspergillus fumigatus.
 XX XX WO200286090-A2.
 PN XX 31-OCT-2002.
 PD XX
 XX XX 23-APR-2002; 2002WO-US013142.
 PF XX
 XX XX 23-APR-2001; 2001US-0285697P.
 PR XX 27-APR-2001; 2001US-0287066P.
 PR XX 05-JUN-2001; 2001US-0295890P.
 PR XX 09-JUL-2001; 2001US-0303899P.
 PR XX 31-AUG-2001; 2001US-0316362P.
 XX XX (ELIT-) ELITRA PHARM INC.
 PA XX
 XX XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 PI XX WPI; 2003-093124/08.
 DR XX
 XX XX New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX XX
 PS Disclosure; Page; 175pp; English.
 XX XX
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX

SQ Sequence 2232 BP; 473 A; 605 C; 545 G; 609 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.88e-11 Length: 2232
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x ABT20789 (1-2232)

QY 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 DB 625 TACATGGTGGCATGGGTGGGTATGTTTCATTACGAACCTGATCCCTCCGACGTT 684

QY 223 Phe 223

DB 685 TTT 687

RESULT 9

ADB69900
 ID ADB69900 standard; DNA; 2456 BP.

AC ADB69900;

DT 04-DEC-2003 (first entry)

DE C. neoformans open reading frame SEQ ID NO:2305.

KW ds; gene; fungicide; gene therapy; infection.

OS Cryptococcus neoformans.

PN WO2003052076-A2.

PD 26-JUN-2003.

PF 17-DEC-2002; 2002WO-US040225.

PR 17-DEC-2001; 2001US-0341261P.

PA (ELIT-) ELITRA PHARM INC.

PI Zamudio C, Eroshkin AM;

DR WPI; 2003-533017/50.

DR P-PSDB; ADB70261.

PT New nucleic acid, useful for preparing a composition for treating an infection caused by *Cryptococcus neoformans*.

PS Claim 2; SEQ ID NO 2305; 136pp; English.

CC The invention relates to a novel purified or isolated *Cryptococcus neoformans* nucleic acid molecule comprising a sequence encoding a polypeptide comprising a sequence not given in the specification. A polynucleotide of the invention has fungicide activity, and may have a use in gene therapy. The nucleic acid is useful for preparing a composition for treating an infection caused by *Cryptococcus neoformans*. The present sequence represents a *C. neoformans* sequence of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 2456 BP; 508 A; 678 C; 580 G; 700 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.47e-11 Length: 2456
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0

DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADB69900 (1-2456)

QY 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171

DB 637 ATTGTACCGGATACATCTCTCGATCTGTGCGCGGTCTTATGACACGAAGCCATTGCC 696

QY 172 Ile 172

DB 697 ATC 699

RESULT 10

ABT20191

ID ABT20191 standard; DNA; 2603 BP.

XX AC ABT20191;

DT 16-APR-2003 (first entry)

DE *Aspergillus fumigatus* essential gene #2549.

KW Fungicide; cytostatic; essential gene; *Aspergillus fumigatus*; infection; cancer; contamination; biofilm; antibody; immune response; ds.

OS *Aspergillus fumigatus*.

PN WO200286090-A2.

PD 31-OCT-2002.

PF 23-APR-2002; 2002WO-US013142.

PR 23-APR-2001; 2001US-0285697P.

PR 27-APR-2001; 2001US-0287066P.

PR 05-JUN-2001; 2001US-0295890P.

PR 09-JUL-2001; 2001US-0303899P.

PR 31-AUG-2001; 2001US-0316362P.

PA (ELIT-) ELITRA PHARM INC.

PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

DR WPI; 2003-093124/08.

PT New purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*, useful for treating or preventing infections by *A. fumigatus*, or for treating a non-infectious disease in a subject e.g. cancer.

PS Disclosure; Page: 175pp; English.

CC The invention relates to novel purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as *A. fumigatus*, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or contain contamination of an object by *A. fumigatus*, or to prevent or inhibit formation on a surface of a biofilm comprising *A. fumigatus*. The polynucleotides are useful for expressing recombinant protein for characterization, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of *A. fumigatus* to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparing with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, as a marker for

CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention

SQ Sequence 2603 BP; 565 A; 698 C; 620 G; 720 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 6,81e-11 Length: 2603
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x ABB20191 (1-2603)

QY 203 TyrMetValSerIaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 DB 877 TACATGTGTGGCATGGGTGGGTATGTTTCATTACGACCTGATCCCTGCACGTT 936

QY 223 Phe 223
 DB 937 TTT 939

RESULT 11

ADB69539
 ID ADB69539 standard; DNA; 2882 BP.

XX AC ADB69539;
 XX DT 04-DEC-2003 (first entry)
 XX DE C. neoformans coding sequence with introns SEQ ID NO:1305.

XX ds; gene; fungicide; gene therapy; infection.

XX OS Cryptococcus neoformans.

XX PN WO2003052076-A2.

XX PD 26-JUN-2003.

XX PF 17-DEC-2002; 2002WO-US040225.
 XX PR 17-DEC-2001; 2001US-0341261P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Zamudio C, Eroshkin AM;

XX DR WPI; 2003-533017/50.

XX DR P-PSDB; ADB70261.

XX New nucleic acid, useful for preparing a composition for treating an
 PT infection caused by *Cryptococcus neoformans*.
 XX Claim 3; SEQ ID NO 1305; 136pp; English.

XX The invention relates to a novel purified or isolated *Cryptococcus*
 CC neoformans nucleic acid molecule comprising a sequence encoding a
 CC polypeptide comprising a sequence not given in the specification. A
 CC polynucleotide of the invention has fungicide activity, and may have a
 CC use in gene therapy. The nucleic acid is useful for preparing a
 CC composition for treating an infection caused by *Cryptococcus neoformans*.
 CC The present sequence represents a C. neoformans sequence of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 2882 BP; 615 A; 759 C; 668 G; 840 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7,51e-11 Length: 2882
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADB69539 (1-2882)

QY 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171
 DB 796 ATTGTACCCGGATACATCTCTCGATCTGTCCCGGTCCTTATGACACGAGCCATTGCC 855
 QY 172 Ile 172
 DB 856 ATC 858

RESULT 12

ABT17781
 ID ABT17781 standard; DNA; 3969 BP.

XX AC ABT17781;

XX DT 16-APR-2003 (first entry)

XX DE *Aspergillus fumigatus* essential gene #139.

XX Fungicide; cytostatic; essential gene; *Aspergillus fumigatus*; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.

XX OS *Aspergillus fumigatus*.

XX PN WO200286090-A2.

XX PD 31-OCT-2002.

XX PF 23-APR-2002; 2002WO-US013142.

XX PR 23-APR-2001; 2001US-0285697P.

XX PR 27-APR-2001; 2001US-0287066P.

XX PR 05-JUN-2001; 2001US-0285890P.

XX PR 09-JUL-2001; 2001US-0303899P.

XX PR 31-AUG-2001; 2001US-0316362P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

XX DR WPI; 2003-093124/08.

XX New purified or isolated nucleic acids of essential genes of *Aspergillus*
 PT *fumigatus*, useful for treating or preventing infections by *A. fumigatus*,
 PT or for treating a non-infectious disease in a subject e.g. cancer.

XX Disclosure; Page; 175pp; English.

XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for
 CC expressing recombinant protein for characterization, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*
 CC *fumigatus* to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune

CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention
 XX Sequence 3969 BP; 964 A; 996 C; 953 G; 1056 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.02e-10 Length: 3969
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x ABT17781 (1-3969)

Qy 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 Db 1514 TACATGGTGTGGCATGGGTGGGTATGCTTTCATTACGACCTGATCCCCCTGCACGTT 1573

Qy 223 Phe 223
 Db 1574 TTT 1576

RESULT 13
 ABT19595
 ID ABT19595 standard; DNA; 4503 BP.
 XX ABT19595;
 AC ABT19595;

DT 16-APR-2003 (first entry)
 XX *Aspergillus fumigatus* essential gene #1953.

DE Fungicide; cytostatic; essential gene; *Aspergillus fumigatus*; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX *Aspergillus fumigatus*.

OS *Aspergillus fumigatus*.

PN WO200286090-A2.

XX 31-OCT-2002.

PF 23-APR-2002; 2002WO-US013142.

XX 23-APR-2001; 2001US-0285697P.

PR 27-APR-2001; 2001US-0287066P.

PR 05-JUN-2001; 2001US-0295890P.

PR 09-JUL-2001; 2001US-0303899P.

PR 31-AUG-2001; 2001US-0316362P.

XX (ELIT-) ELITRA PHARM INC.

PA Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

XX WPI; 2003-093124/08.

DR New purified or isolated nucleic acids of essential genes of *Aspergillus*

PT *fumigatus*, useful for treating or preventing infections by *A. fumigatus*,
 PT or for treating a non-infectious disease in a subject e.g. cancer.

XX Disclosure; Page; 175pp; English.

XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a

CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*
 CC *fumigatus* to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention

SQ Sequence 4603 BP; 1121 A; 1129 C; 1133 G; 1220 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.17e-10 Length: 4503
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x ABT19595 (1-4503)

Qy 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 Db 1877 TACATGGTGTGGCATGGGTGGGTATGCTTTCATTACGACCTGATCCCCCTGCACGTT 1936

Qy 223 Phe 223
 Db 1937 TTT 1939

RESULT 14
 ADB69178
 ID ADB69178 standard; DNA; 4738 BP.

XX ADB69178;

XX 04-DEC-2003 (first entry)

DE C. neoformans genomic DNA sequence SEQ ID NO:305.

XX ds; gene; fungicide; gene therapy; infection.

XX *Cryptococcus neoformans*.

XX WO20003052076-A2.

XX 26-JUN-2003.

XX 17-DEC-2002; 2002WO-US040225.

XX 17-DEC-2001; 2001US-0341261P.

XX (ELIT-) ELITRA PHARM INC.

XX Zamudio C, Eroshkin AM;

XX WPI; 2003-533017/50.

XX P-PADB; ADB70261.

XX New nucleic acid, useful for preparing a composition for treating an

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 139.895 Seconds
(without alignments)
3648.073 Million cell updates/sec

Title: US-10-028-384-6
Perfect score: 718
Sequence: 1 WGSRSCLSVFQTIKLVI.....TRSTKTRRSIKRPELGLRV 718

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0
Fgapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spoci/US10028384/runat_14122004_131646_6816/app.query.fasta_1.3740
-DB=Issued Patents NA -OFFMT=fastap -SUFFIX=Oligo.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Oligo -TRANS=human40.ccd
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US10028384 @CNG 1 1 259 @runat_14122004_131646_6816 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGPOP=60 -XGAPEXT=60 -FCAPOP=5
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOF=6 -DELEXT=7

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*

6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	631	87.9	2157	US-09-614-221A-318	Sequence 318, App
2	30	4.2	857	US-09-248-796A-3089	Sequence 3089, Ap
3	26	3.6	594	US-09-248-796A-3090	Sequence 3090, Ap
4	20	2.8	900	US-09-270-767-12331	Sequence 12331, A
5	17	2.4	1285	US-09-270-767-11648	Sequence 11648, A
6	16	2.2	225	US-09-513-999C-21090	Sequence 21090, A
7	16	2.2	580	US-09-270-767-1318	Sequence 1318, Ap
8	16	2.2	560	US-09-270-767-16600	Sequence 16600, A
9	14	1.9	503	US-09-270-767-27262	Sequence 27262, A
10	14	1.9	616	US-09-328-111-332	Sequence 332, App
11	13	1.8	302	US-09-313-294A-6869	Sequence 6869, Ap
12	12	1.7	265	US-09-313-294A-910	Sequence 910, App

ALIGNMENTS

RESULT 1

US-09-614-221A-318
; Sequence 318, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunananda, Balasuliojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614, 221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 318
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-318

Alignment Scores:	Pred. No.:	Score:	Length:
Best Local Similarity:	0	631.00	2157
Query Match:	87.92%	Conservative:	717
DB:	87.88%	Mismatches:	0
	4	Indels:	2
		Gaps:	0

US-10-028-384-6 (1-718) x US-09-614-221A-318 (1-2157)

QY 1 MetGlyGerAspArgSerCysValLeuSerValPheGlnThrIleuLysLeuValIle 20
Db 1 ATGGGATCCGACCGGTGCTGTGTTTGTCTGTGTTTTCAGACCATCTCCTCAAGCTCGTCATC 60

QY 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleIleLysPheGlu 40
DB 61 TTCTGGTGGCGCATTTTGGGGCTGCCATATCATCAGTTTGTTCAGTCATCAAAATTGAG 120
QY 41 SerIleIleHisGluPheAspProTTPheAsnTyrArgAlaThrLysIleLysValLeuVal 60
DB 121 TCTATATCCATGATTCAGCCCTGGTTCATATATAGGGCTACCAAAATATCTGTCAAC 180
QY 61 AsnSerPheTyrLysPheLeuAsnTTPheAspAspArgThrTTPyrProLeuGlyArg 80
DB 181 AATTCGTTTTCAGATTTTGAACCTGGTTTCAGCCAGTACCTGGTACCCCTCGGAAGG 240
QY 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrPheAla 100
DB 241 GTTACTGGAGGACTTTATATCTCTGGTTTGTATGACGACTAGTGGTTCATCTGGCAGCC 300
QY 101 LeuArgAsnTTPyrLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
DB 301 CTGGCCAACTGGTTGGCTTCCCATTCATCATCAGAAACGTTTGTGTATTTGGCCCA 360
QY 121 LeuPheSerGlyValThrAlaThrAlaThrTyrGluPheThrLysGluIleLysAspAla 140
DB 361 CTATTTCTGGGGTCACCGCTGGCGACTTACGATTTACGAAAGAGATTAAGATGCC 420
QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
DB 421 AGCGCTGGGCTTTGGCTGGCTTTTATAGCCATTTGCCCGGTTATATATCTAGATCA 480
QY 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
DB 481 GTGGCGGGTCTACGATAATAGGCGCATTCGCCATTCACACTAATATGTCATTTTCATG 540
QY 181 PheThrIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
DB 541 TTTTGGATTAGGCGCCAAAGACTGGCTCTATCATGCGCAACAGTGTGCGCTTTATTC 600
QY 201 TyrPheTyrMetValSerAlaThrPyrGlyTyrValPheIleThrAsnLeuIleProLeu 220
DB 601 TACTTCTCATGTTGGTGGCTGGGTGGATAGTGTTCATCACCACATTCATGCCACTC 660
QY 221 HisValPheLeuLeuIleLeuMetClyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
DB 661 CATGCTTTTCTGCTATTTGATGGCGAGATATCGTCCAACTGTATTCGCCCTACACC 720
QY 241 ThrTTPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
DB 721 ACTTGGTACGCTATTGGAACTTTGTCATCCATCGAGATCCCATTTGTGGTTTCTTACCT 780
QY 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
DB 781 ATCAGTCTACGACCACTGGCCGCAATGGGTGGTTTTCGTTTGTATTCAGATTTGCGCC 840
QY 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
DB 841 TTCGGTGTACTTCGTGAAGGGCCAAATCAGCACAGCTAAAGTTTAAAGTCTCATGTGTT 900
QY 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
DB 901 TCTCTGTTTTCATCTTGGTCTTGGTGGTGGACTTTCGCCCTTTCCTATATATGGGG 960
QY 321 LeuIleAlaProTTPyrThrGlyArgPheTyrSerLeuTTPasPheThrAsnTyrAlaLysIle 340
DB 961 TTGATTTGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCAACTACGCAAGATC 1020
QY 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTTProAlaPhePhe 360
DB 1021 CACATTCCTATCATTCCTCGTTCCGAACATCACCCTGTTTCGTGGCCCGCTTTCCTC 1080
QY 361 PheAspThrHisPheLeuIleThrPheProAlaGlyValPheLeuLeuPheLeuAsp 380
DB 1081 TTTGTATCCCACTTTTGTATCTGGCTATTCGCCGGGTGTATTCCTACTATTCCTCGAC 1140
QY 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuLysSerTyrPheAlaGly 400

DB 1141 TTCAAAGACGACACAGCTTTTGTTCATCGCTTACTCCGTTCTGTGTTCGTACTTTGCCGGT 1200
QY 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
DB 1201 GTTATGGTTAGATTGATGTTCACCTTTCAGCCAGCTCATCTGTGTGTCGCGCGCTCGCA 1260
QY 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
DB 1261 TTGTCCAGATATTTGACATCTACCTGGATTTCAGGACAAAGTACGCCAAATACGCCATC 1320
QY 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
DB 1321 AAACCTCGCGCACTACTGCCAAATTTGATTTTCCGGATCATTCATCTTTTATTGTAT 1380
QY 461 LeuPheValPheHisSerThrTTPyrValThrArgThrAlaTyrSerSerProSerValVal 480
DB 1381 CTTTTCGCTCTTCATCTTACTTGGGTAAACAAGAACTGCATACTCTTCTCTCTGTGT 1440
QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
DB 1441 TTGCCATCAAAACCCAGATGGTAAATTTGGCGTTGATCGACGACTTCAGGGAAGCGTAC 1500
QY 501 TyrTTPLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTTPyrAspTyrGly 520
DB 1501 TATTTGGTTAAGAAATGAATCTGTATGAGACAGTAAAGTTTGCAGCGTGGTGGGATTACGGT 1560
QY 521 TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTTPAspAsn 540
DB 1561 TACCAAAATGGTGGCATGGCAGACAGAACCACTTTTGTTCGATAACAACACGTGGAACAAT 1620
QY 541 ThrHisIleAlaIleValGlyValAlaMetAlaSerProGluGluLysSerTyrGluIle 560
DB 1621 ACTCATCTCGCATCTGTGTGGTAAAGCCATGCTTCCCTCGAAGAAATCTTACGAAT 1680
QY 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
DB 1681 CTAAAGAGCATGATGTCGATTTATGTCATCTTGTGTCTTGTCTTAATTTGGTTTGT 1740
QY 581 GlyAspAspIleAsnLysPheLeuTTPyrMetIleArgIleSerGluGlyIleTTPProGlu 600
DB 1741 GGTGATGATCATCAAAATTTCTTGTGATGATTCAGATTCAGAGGAGATCTGGCCAGAA 1800
QY 601 GluIleLysGluArgTyrPheTyrThrAlaGluGlyGlyTyrArgValAspAlaArgAl 620
DB 1801 GAGATAAAGACGCTGA-TTTCTATACCGCAGAGGAGAAATACAGAGTAGATGCAAGGCG 1859
QY 620 aSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLe 640
DB 1860 TTCTGAGACCATGAGGAACTCGCTACTTTACAGATGTCCTACAGAGATTTCCCAAT 1919
QY 640 uPheAsnGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal 660
DB 1920 ATTCAATGGTGGCCAAAGCACTGACAGAGTGGGTCAACAAATGATCACCATTAGACGT 1979
QY 660 lProProLeuAspTyrPheAspGluValPheThrSerGluAsnTTPyrMetValArgIleTyr 680
DB 1980 CCCACCATTAGACTTCTGACGAGATTTTACTTCCGAAAACTGGATGGTTAGATATA 2039
QY 680 rGlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
DB 2040 TCAATTTGAAGAGGATGATGCCAAGTAGAACTTTGAGGAGCGTTGGTGAGTTAACCCAG 2099
QY 700 gSerSerThrLysThrArgArgSerIleLysArgTTProGluLeuGlyLeuArgVal 718
DB 2100 GTCCTCTACGAAACCAAGAGTCCATAAAGAGACCTGAATTAGGCTTGAGAGTC 2154

RESULT 2

US-09-248-796A-3089

Sequence 3089, Application US/09248796A

Patent No. 6747137

GENERAL INFORMATION:

APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3089
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (25)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown

US-09-248-796A-3089

Alignment Scores:
Pred. No.: 1,61e-21 Length: 967
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4,18% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-248-796A-3089 (1-867)

QY 515 AlaTTPtPAspTyrGlyTyrGlnleGlyMetAlaAspArgThrThrLeuValasp 534

Db 244 GCCTGGTGGATTATGGTTATCAATCGGGGTATGCTGATAGACACACATGGTAT 303

QY 535 AsnAsnThrTyrAsnAsnThrHisleAla 544

Db 304 AACATACATGATGAATACACACATATGCC 333

RESULT 3

US-09-248-796A-3090

; Sequence 3090.. Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 3090

; LENGTH: 594

; TYPE: DNA

; ORGANISM: Candida albicans

US-09-248-796A-3090

Alignment Scores:

Pred. No.: 1.6e-17 Length: 594
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3,62% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-248-796A-3090 (1-594)

QY 66 PheLeuAsnTrpPheAspArgThrTyrProLeuGlyArgValThrGlyThr 85

Db 241 TTTTGAATGGTTTATGATAGACATGGTACCACTGGAGAGATCACTGGTGTACT 300

QY 86 LeuTyrProGlyLeuMet 91
Db 301 TTATATCCGGTTTAAAG 318

RESULT 4

US-09-270-767-12331

; Sequence 12331.. Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12331

; LENGTH: 900

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

US-09-270-767-12331

Alignment Scores:

Pred. No.: 4.13e-11 Length: 900
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2,79% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-270-767-12331 (1-900)

QY 149 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168

Db 532 TTCATCGCCATCGTGGCTGCTACATCAGTAGTGGCTGCTGATACGATACGATACGAG 591

RESULT 5

US-09-270-767-11648

; Sequence 11648.. Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 11648

; LENGTH: 1386

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

US-09-270-767-11648

Alignment Scores:

Pred. No.: 8.22e-08 Length: 1386
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2,37% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-270-767-11648 (1-1386)

QY 152 ILeValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168

Db 876 ATCGTCCCGGGTATATCTCTCGATCGGTGGCGGATCGTACGACATGA 926

RESULT 6

US-09-513-999C-21090

; Sequence 21090.. Application US/09513999C

; Patent No. 6783961

```

; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; PATENT NO. 6783961
; FILE REFERENCE: 59. US2.REG
; CURRENT APPLICATION NUMBER: US/09/513.999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 21090
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-21090

Alignment Scores:
Pred. No.: 1.7e-07 Length: 245
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.23% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-513-999C-21090 (1-245)

QY 149 PhelealalValProGlyTyrIleSerArgSerValalaGlySer 164
DB 60 TTTATTGCTATTGTACCAAGCTACATATCTCGGTGATGCTAGTGGATCC-107

RESULT 7
US-09-270-767-1318/c
; Sequence 1318, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1318
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1318

Alignment Scores:
Pred. No.: 3.76e-07 Length: 560
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.23% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-270-767-1318 (1-560)

QY 397 TyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProVal 412
DB 402 TACTTCGCTGGTGTGATGGTGGTTTGATGTTGACCTCAGCCGGTG 355

RESULT 8
US-09-270-767-16600/c
; Sequence 16600, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094

```

```

; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16600
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16600

Alignment Scores:
Pred. No.: 3.76e-07 Length: 560
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.23% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-270-767-16600 (1-560)

QY 397 TyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProVal 412
DB 402 TACTTCGCTGGTGTGATGGTGGTTTGATGTTGACCTCAGCCGGTG 355

RESULT 9
US-09-270-767-27262
; Sequence 27262, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27262
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-27262

Alignment Scores:
Pred. No.: 4.07e-05 Length: 503
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.95% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-270-767-27262 (1-503)

QY 155 GlyTyrIleSerArgSerValalaGlySerTyrAspGlu 168
DB 2 GGGTATATCTCTCGATCCGTGGCGGATCGTACGACATGAA 43

RESULT 10
US-09-328-111-332
; Sequence 332, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert

```


; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 332
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(616)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-332

Alignment Scores:
Pred. No.: 4.94e-05 Length: 616
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.95% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-6 (1-718) x US-09-328-111-332 (1-616)

Qy 201 TyPheTyrMetValSerAlaTrpGlyTyrValPheille 214
Db 149 TATTTCATATGCTCTCTGTTGGGTGTTATGTTATTC 190

RESULT 11

US-09-313-294A-6869
; Sequence 6869, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 6869
; LENGTH: 302
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700352557H1/
US-09-313-294A-6869

Alignment Scores:
Pred. No.: 0.000272 Length: 302
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.81% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-313-294A-6869 (1-302)

Qy 341 HisIleProIleAlaSerValSerGluHisGlnPro 353
Db 11 CATATACCAATCATCGCATCTGTTTCTGAGCATGCCA 49

RESULT 12

US-09-313-294A-910
; Sequence 910, Application US/09313294A

; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 910
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700550053H1
US-09-313-294A-910

Alignment Scores:
Pred. No.: 0.00263 Length: 265
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.67% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-313-294A-910 (1-265)

Qy 70 PheAspArgThrTrpTyrProLeuGlyArgVal 81
Db 3 TTTGATGATAGGACATGTTACCCCTTGCCGTGTG 38

RESULT 13

US-09-313-294A-906
; Sequence 906, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 906
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700550043H1
US-09-313-294A-906

Alignment Scores:
Pred. No.: 0.00267 Length: 269
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.67% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-313-294A-906 (1-269)

Qy 70 PheAspArgThrTrpTyrProLeuGlyArgVal 81
Db 3 TTTGATGATAGGACATGTTACCCCTTGCCGTGTG 38

RESULT 14

US-09-313-294A-4442

us-10-028-384-6.Oligo.rni

Thu Dec 16 16:25:08 2004

Db 256 ATCCCTATATTGCTAGTGTGAGGACCAACCA 291
 Search completed: December 14, 2004, 23:43:44
 Job time : 148.895 secs

; Sequence 4442, Application US/09313294A
 ; Patent No. 6476212
 ; GENERAL INFORMATION:
 ; APPLICANT: Lalgudi, Raghunath V.
 ; APPLICANT: Ito, Laura Y.
 ; APPLICANT: Sherman, Bradley K.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
 ; FILE REFERENCE: PL-0017 US
 ; CURRENT APPLICATION NUMBER: US/09/313,294A
 ; CURRENT FILING DATE: 1999-05-14
 ; NUMBER OF SEQ ID NOS: 7600
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 4442
 ; LENGTH: 292
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. 6476212 700348434H1
 ; NAME/KEY: unsure
 ; LOCATION: 24
 ; OTHER INFORMATION: a, t, c, g, or other
 US-09-313-294A-4442

Alignment Scores:
 Pred. No.: 0.00289 Length: 292
 Score: 12.00 Matches: 12
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.67% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-313-294A-4442 (1-292)

QY 533 ValAspAsnAsnThrTriPAsnAsnThrHisIleAla 544
 DB 150 GTTGATACATACATGGAATTAACACACACATAGCT 185

RESULT 15

US-09-313-294A-1303
 ; Sequence 1303, Application US/09313294A
 ; Patent No. 6476212
 ; GENERAL INFORMATION:
 ; APPLICANT: Lalgudi, Raghunath V.
 ; APPLICANT: Ito, Laura Y.
 ; APPLICANT: Sherman, Bradley K.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
 ; FILE REFERENCE: PL-0017 US
 ; CURRENT APPLICATION NUMBER: US/09/313,294A
 ; CURRENT FILING DATE: 1999-05-14
 ; NUMBER OF SEQ ID NOS: 7600
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 1303
 ; LENGTH: 299
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. 6476212 700550709H1
 US-09-313-294A-1303

Alignment Scores:
 Pred. No.: 0.00295 Length: 299
 Score: 12.00 Matches: 12
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.67% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-313-294A-1303 (1-299)

QY 342 IleProIleIleAlaSerValSerGluHisIlePro 353

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 23:35:31 ; Search time 903.972 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 718

Sequence: 1 MGSRSCLSVFQTLKLV.....TRSTKTRRSIKPELGLRV 718

Scoring table:

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8161859

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n: model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131648_6892/app_query.fasta_1.3740
-DB=Published Applications NA -OFT=fastap -SUFFIX=oligo.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -LOCALIGN=200 -THR SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10028384@cgn_1_1_2124@runat_14122004_131648_6892
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	718	100.0	2733	15	US-10-028-384-5
2	631	87.9	2157	9	Sequence 5, Appli
3	631	87.9	2157	18	Sequence 387, App
4	32	4.5	2256	15	Sequence 318, App
5	21	2.9	1848	15	Sequence 6323, Ap
6	21	2.9	1969	15	Sequence 2139, Ap
7	21	2.9	2232	15	Sequence 1139, Ap
8	21	2.9	2466	16	Sequence 7139, Ap
9	21	2.9	2603	15	Sequence 2305, Ap
10	21	2.9	2882	16	Sequence 6139, Ap
11	21	2.9	3969	15	Sequence 1305, Ap
12	21	2.9	4603	15	Sequence 139, App
13	21	2.9	4738	16	Sequence 5139, App
14	20	2.8	2417	15	Sequence 305, App
15	19	2.6	2681	16	Sequence 7, Appli
16	19	2.6	2839	18	Sequence 14408, A
17	17	2.4	387	16	Sequence 150745,
18	17	2.4	393	16	Sequence 173, App
19	17	2.4	558	14	Sequence 323, App
20	17	2.4	1209	15	Sequence 433, App
21	17	2.4	1209	15	Sequence 310, App
22	17	2.4	1543	10	Sequence 412, App
23	17	2.4	1543	10	Sequence 133, App
24	17	2.4	1543	10	Sequence 133, App
25	17	2.4	1543	11	Sequence 133, App
26	17	2.4	1543	16	Sequence 133, App
27	17	2.4	2481	15	Sequence 133, App
28	17	2.4	2710	15	Sequence 1, Appli
29	16	2.2	154	10	Sequence 3, Appli
30	16	2.2	474	14	Sequence 450, App
31	16	2.2	485	10	Sequence 454, App
32	16	2.2	500	9	Sequence 11283, A
33	16	2.2	1114	16	Sequence 1643, Ap
34	16	2.2	2472	15	Sequence 629, App
35	16	2.2	2472	15	Sequence 112, App
36	16	2.2	2472	15	Sequence 11, Appl
37	16	2.2	2472	16	Sequence 742, App
38	16	2.2	3046	18	Sequence 95, Appl
39	16	2.2	3093	18	Sequence 92, Appl
40	16	2.2	3094	15	Sequence 9, Appli
41	16	2.2	5404	18	Sequence 99, Appl
42	16	2.2	5827	18	Sequence 97, Appl
43	16	2.2	54810	18	Sequence 91, Appl
44	16	2.2	103375	18	Sequence 94, Appl
45	15	2.1	528	18	Sequence 424, App

ALIGNMENTS

RESULT 1
US-10-028-384-5
; Sequence 5, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Kevin
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2733
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: D28952
; DATABASE ENTRY DATE: 1999-02-07

RELEVANT RESIDUES: (1) .. (2733)

US-10-028-384-5

Alignment Scores:

Pred. No.: 0 Length: 2733
 Score: 718.00 Matches: 718
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 15

US-10-028-384-6 (1-718) x US-10-028-384-5 (1-2733)

QY	1	MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValile	20
DB	219	ATGGATCCGACGGTGTGTGTTTGTCTGTGTTCAGCCATCTCAAGCTCGTCATC	278
QY	21	PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu	40
DB	279	TTCTGGCGGATTTTGGGGCTGCATATCATCACGCTTTGTTTGCAGTCATCAAAATTTGAG	338
QY	41	SerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrLysTyrLeuValAsn	60
DB	339	TCATATATCCATGATTCGACCCCTGTTCATATTATAGGGCTACCAATATCTCGTCAC	398
QY	61	AsnSerPheTyrLysPheLeuAsnTyrPheAspAspArgThrTyrTyrProLeuGlyArg	80
DB	399	AAITCGTPTTACAAAGTTTGTAACTGGTTTGACGACCGTACCTGGTACCCCTCGGAAG	458
QY	81	ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla	100
DB	459	GTTACTGGGGGACTTATATCTCTGTGTGTGACGACTAGTGGCTCATCTGGCACGCC	518
QY	101	LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro	120
DB	519	CTGGCAACTGGTGGCTTGGCCATTGACATGACAGAAAGTTGTGTGTATTTCGGCCA	578
QY	121	LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla	140
DB	579	CTATTTCTGGGTACCCGCTGGGCGACTTACGAAATTTACGAAGAGATTAAAGATGCC	638
QY	141	SerAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuMetValThrPheMet	160
DB	639	AGCCTCGGGTGTGGCTGTGTGTATAGCAATGTCTCCCGGTTATATATCTAGATCA	698
QY	161	ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuMetValThrPheMet	180
DB	699	GTGGCGGGTCTTACGATATAGGCCATTGCCATTACACTATTATATGGTCACTTTCATG	758
QY	181	PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe	200
DB	759	TTTGGATTAAGCCCAAGACTGGCTCTATCATGACGCAACGTGTGCGAGCTTTATTC	818
QY	201	TyrPheTyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeu	220
DB	819	TACTTCTACATGGTGTGGCTGGGGTGGATACGTGTTCATCCCAACTTGATCCCACTC	878
QY	221	HisValPheLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr	240
DB	879	CATGTCTTTTGTGTGATTGTGGGCGAGATATTCGTCAAACTGTATTCTGGCTACACC	938
QY	241	ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro	260
DB	939	ACTTGTGTACCTATTGGAACTGTTCATCCATCCAGATCCCAATTTGTGGTTTCTACCT	998
QY	261	IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla	280
DB	999	ATCAGTCTAACGACCAATGGCCGCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1058
QY	281	PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal	300
DB	1059	TTGGGTGACTTCGTGAAGGGCCAAATCAGCACAGCTTAAGTTTAAAGTTCATCATGTGTT	1118

QY	301	SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly	320
DB	1119	TTCTGTGTTTGTGATCTTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1178
QY	321	LeuIleAlaProTyrThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle	340
DB	1179	TTGATTTGCCCTTGGACTGGTAGATTATTTCGTATGGGATACCACTACGCAAGATC	1238
QY	341	HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProIlePhePhe	360
DB	1239	CACATTTCTATCATTTGCTCGCTTCCGAACATCAACCCGTTTCGTGGCCGCTTTCTTC	1298
QY	361	PheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp	380
DB	1299	TTTGATACCCACTTTTGTATCTGGCTATTCCCGCGGTGTATTCTTCTACTATTCTTCGAC	1358
QY	381	LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly	400
DB	1359	TTGAAAGACGACGACGCTTTTGTCTCATGCTTACTCCGTTCTGTGTCTGTACTTTTGGCGGT	1418
QY	401	ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla	420
DB	1419	GTTATGGTTAGATTGATGTTGACTTTTGACACCACTCATCTGTGTGTGTGTGTGTGTGTGT	1478
QY	421	LeuSerLysIlePheAspIleTyrIleAspPheLysThrSerAspArgLysTyrAlaIle	440
DB	1479	TTGTCCAAAGATATTGTACATCTACTCTGATTTCAAGACAAAGTACCCGCAAAATACGCCATC	1538
QY	441	LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr	460
DB	1539	AAACCTCGGCGACTTACTGGCCAAATTTGATTTGTTTCGGATCATTTCACTTTTATTGTAT	1598
QY	461	LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerValVal	480
DB	1599	CTTTTCTGTCTTCCATTTCTACTTGGGTAAACAGAACTGCATCTCTCTCTCTCTCTCTCTCT	1658
QY	481	LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr	500
DB	1659	TTGCCATCACAAACCCAGATGTTAAATTTGGCGTTGTATCGCGACTTCAGGGAAGCGTAC	1718
QY	501	TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaIleTrpTrpAspTyrGly	520
DB	1719	TATTGGTTAAAGATGAATCTGATGAGGACAGTAAGGTTCAGCGTGTGGGATACGGT	1778
QY	521	TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn	540
DB	1779	TACCAATTTGGTGGCATGGCAGACAGAACCCACTTTTACTCGATTAACACACGTCGAACAT	1938
QY	541	ThrHisIleAlaIleValGlyAlaValAlaSerProGluGluLysSerTyrGluIle	560
DB	1839	ACTCATCTGGCCATCGTTGGTAAAGCCATGGCTTCCCTCGAAGAGAAATCTTACGAAAT	1898
QY	561	LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly	580
DB	1899	CTAAAGAGCATGATGTGATTTATGTTCTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT	1958
QY	581	GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu	600
DB	1959	GGTGTGTGATCAACAAATTTCTGTGATGATCAAGAAATTAGCGAGGAAATCTCGCCAGAA	2018
QY	601	GluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAla	620
DB	2019	GAGATAAAGAGCGTTATTTCTATACCGCAGAGGAGAGATACAGATAGATGTCAGGGCT	2078
QY	621	SerGluThrMetArgAsnSerLeuLeuTyrIleMetSerTyrLysAspPheProGlnLeu	640
DB	2079	TCTGAGACCATGAGGAATCTGCTACTTTACAGATGTCTTACAAAGATTTTCCACAAATA	2138
QY	641	PheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal	660
DB	2139	TTCAATGTGGCCAAAGCCACTGACAGAGTGCCTCAACAAATGATCACACCATTAGACGTC	2198
QY	661	ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyr	680

Db 2199 CCACATTAGACTACTTCGACGAAGTTTATCTTCGAAAACCTGGATGGTTAGATATAT 2358
 Qy 681 GlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
 Db 2259 CAATTGAGAGAGATGATCCCAAGGTAGAACTTTGAGGGACGTTGGTGAAGTTAACGAGG 2318
 Qy 701 SerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal 718
 Db 2319 TCTTCTACGAAACACGAGAGGTCATTAAGAGACCTGAATTAGGCTTGAGAGTC 2372

RESULT 2

US-09-801-368-387
 ; Sequence 387, Application US/09801368
 ; Patent No. US20020128250A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Busby, Robert
 ; APPLICANT: Cali, Brian
 ; APPLICANT: Hecht, Peter
 ; APPLICANT: Holtzman, Doug
 ; APPLICANT: Madden, Kevin
 ; APPLICANT: Maxon, Mary
 ; APPLICANT: Milne, Todd
 ; APPLICANT: Royer, John
 ; APPLICANT: Salama, Sofie
 ; APPLICANT: Sherman, Amir
 ; APPLICANT: Silva, Jeff
 ; APPLICANT: Summers, Eric
 ; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
 ; FILE REFERENCE: 109272.147
 ; CURRENT APPLICATION NUMBER: US/09/801,368
 ; CURRENT FILING DATE: 2001-03-07
 ; PRIOR APPLICATION NUMBER: US 09/487,558
 ; PRIOR FILING DATE: 2000-01-19
 ; PRIOR APPLICATION NUMBER: US 60/160,587
 ; PRIOR FILING DATE: 1999-10-20
 ; NUMBER OF SEQ ID NOS: 440
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 387
 ; LENGTH: 2157
 ; TYPE: DNA
 ; ORGANISM: Saccharomyces cerevisiae
 US-09-801-368-387

Alignment Scores:

Pred. No.: 0 Length: 2157
 Score: 631.00 Matches: 717
 Percent Similarity: 99.72% Conservative: 0
 Best Local Similarity: 99.72% Mismatches: 1
 Query Match: 87.88% Indels: 2
 DB: 9 Gaps: 0

US-10-028-384-6 (1-718) x US-09-801-368-387 (1-2157)

Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValile 20
 Db 1 ATGGGATCCGACCGGTCGTGTGTTTGTCTGTGTTTCACACCATCTCAAGCTGTCATC 60
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 61 TTCGTGGCAATTTTGGGGCTGCCATATCATCAGCTTGTGTCAGTCATCAAAATTGAG 120
 Qy 41 SerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrLysTrpLeuValAsn 60
 Db 121 TCTATTATCATCAATTCACACCCCTGGTTCAATTATAGGCTACCAATATCTCGTCAAC 180
 Qy 61 AsnSerPheTyLysPheLeuAsnTrpPheAspArgThrTrpTyProLeuGlyArg 80
 Db 181 AATTCGTTTAAAGFTTTTGAACGTGTTGACGACCGTACCTGTGATCCCTCCGGAAGG 240
 Qy 81 ValThrGlyGlyThrLeuTrpProGlyLeuMetThrThrSerAlaPheIleTrpHisala 100

Db 241 GTTACTGGAGGACCTTTATATCTCTGGTTTGATGACGACCTAGTGGCTTCATCTGGCACGCC 300
 Qy 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 Db 301 CTGCGCAACTGGTGTGGCTTGCCATTGACATCAGAAACGTTGTGTGCTATTTCGGCCA 360
 Qy 121 LeuPheSerGlyValThrAlaTrpAlaThrTyGluPheThrLysGluIleLysAspAla 140
 Db 361 CTATTTTCTGGGGTCAACCGCTTGGGGACCTTACGAATTTACGAAAGAGATTAAAGATGCC 420
 Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyIleSerArgSer 160
 Db 421 AGCGCTGGGCTTTTGGCTGCTGCTTTTATAGCCATTGTCGCCGGTTATATCTAGATCA 480
 Qy 161 ValAlaGlySerTyAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 481 GTGGCGGGGTCTCTAGATAATGAGGCCATTGCCATTACACTATTAAATGGTCACCTTCATG 540
 Qy 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 Db 541 TTTTGGATTAAAGGCCCAAAAGACTGGCTCTATCATGACGACGACGTCGAGCTTTATTC 600
 Qy 201 TyrPheTyMetValSerAlaTrpGlyTyTrpValPheIleThrAsnLeuIleProLeu 220
 Db 601 TACTTCTACATGGTCTCGGCTTGGGGTGGATACGTTTCATCACCACACTTGATCCCATC 660
 Qy 221 HisValPheLeuLeuIleLeuMetGlyArgTyTrpSerSerLysLeuTyTrpSerAlaTyThr 240
 Db 661 CATGCTCTTTTGCATGATTTTGAATGGGACAGATATTCGCCAAACTGATCTGCGCTACAC 720
 Qy 241 ThrTrpTyAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 721 ACTTGGTACGCTATTGGAACTGTTCATCCATCCAGATCCCATTTGTGGTTTCTACCT 780
 Qy 261 IleArgSerAsnAspHisMetAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 781 ATCAGTCTTAAGCCACATGCGCGCATTTGGGTGTTTTCGGTTTGATTCAAGATTGTGCC 840
 Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 Db 841 TTCGTGACTTTCGTGAAGGGCCAAATCAGCACAGCTAAAGTTTAAAGTCATCATCGTT 900
 Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyMetGly 320
 Db 901 TCTCTGTTTTTGTATCTGGTCCITGGTGTGGTCCGACTTCTGCTTGAACCTATATGGG 960
 Qy 321 LeuIleAlaProTrpThrGlyArgPheTyTrpSerLeuTrpAspThrAsnTrpAlaLysIle 340
 Db 961 TTGATTGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCACTACGCAAGATC 1020
 Qy 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProAlaPhePhe 360
 Db 1021 CACATTCTATCATTTGCTCCGTTTCCGAAACATCAACCCGTTTCGTGGCCGCTTTCTTC 1080
 Qy 361 PheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 Db 1081 TTTGATACCCACATTTTGTATCTGGCTATTCGCCCGCGGTATTCCTTACTATTCTCTGAC 1140
 Qy 381 LeuLysAspGluHisValPheValIleAlaTyTrpValLeuCysSerTyTrpPheAlaGly 400
 Db 1141 TTGAAGACGACGACGCTTTTGTGTCATCGCTTACTCCGTTCTGTGTTCTGTTTGGCGGT 1200
 Qy 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
 Db 1201 GTTATGTTAGATTGATTGATTTTGACACCATCATCTGTGTGTCGCCGCGCGCTGCA 1260
 Qy 421 LeuSerLysIlePheAspIleTyTrpLeuAspPheLysThrSerAspArgLysTyAlaIle 440
 Db 1261 TTGTCGAAGATATTGACATCTACCTGGATTTCAGACAGTTCGCCCAANTACGCAATC 1320
 Qy 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyLeuTy 460
 Db 1321 AAACCTGCGGCACCTACTGGCCAAATGATTGTTTCCGGATCATTCATCTTTTATTGTAT 1380

```

QY 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerProSerValVal 480
Db 1381 CTTTTCGCTTCCATTCTACTTGGTAAACAGAACTGCATCTCTCTCTCTCTCTT 1440
QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuLeuAspPheArgGluAlaTyr 500
Db 1441 TTGCATCAACAAACCCAGATGTAATTTGGCTTGCATCGACGACTTCAGGAAACGATAC 1500
QY 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaAlaTyrTrpGly 520
Db 1501 TATTGGTTAAGATGAATCTCTGATGAGGACAGTAAGTTGCAGCGTGGTGGATACGGT 1560
QY 521 TyrGlnThrGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
Db 1561 TACCAATTTGGTGGCATGGGAGAGAGAACCTTTTATGTCGATAACAAACGTTGGAAACAT 1620
QY 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle 560
Db 1621 ACTCACATGCCATCGTTGTTAAAGCATGGCTTCCCTGAAGAGAAATCTTACGAATTT 1680
QY 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
Db 1681 CTAAAGAGAGATGATGTCGATTAATGCTTGGTCATCTTTGGTGGTCTAATTTGGGTTGGT 1740
QY 581 GlyAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyLysIleTrpProGlu 600
Db 1741 GGTGATGACATCAACAAATCTTGTGATGATCAGAAATTAGCGAGGAATCTGGCCAGAA 1800
QY 601 GluIleLysGluArgTyr-PheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAl 620
Db 1801 GAGATAAAGAGCGTGA-TTTCATACCCGAGGAGGAGAAATACAGAGTAGATGCAAGGCG 1859
QY 620 aSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLe 640
Db 1860 TTCTGAGACCATGAGGAATCGTACTTTTACAAGATGCTCTACAAGATTTCCCACAAT 1919
QY 640 uPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa 660
Db 1920 ATTCAATGGTGGCCAGCCACTACAGAGTGGCTCAACAAATGATCACCACCATTAGACGT 1979
QY 660 lProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyr 680
Db 1980 CCCACCATTAGACTACTTCCAGCAAGTTTACTTCCGAAACTGGATGGTGTAGATATA 2039
QY 680 rGlnLeuLysLysAspAspAlaGlnArgThrLeuArgAspValGlyLeuThrArg 700
Db 2040 TCAATTGAAGAGGATGATGCCCAAGGTAGAACTTTGAGGGACGTTGGTGGATTAACGAG 2099
QY 700 gSerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal 718
Db 2100 GTCTTCTACGAAACACAGAGGTCCTATAAGAGACCTGAATTAGGCTTGAGATC 2154

```

RESULT 3

```

US-10-793-639-318
; Sequence 318, Application US/10793639
; Publication No. US20040199940A1
; GENERAL INFORMATION:
; APPLICANT: Kartunandaa, Balasulajini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/10/793,639
; PRIOR FILING DATE: 2004-03-05
; PRIOR APPLICATION NUMBER: US/09/614,221A
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: US 60/142,981
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 318
; LENGTH: 2157

```

```

; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-793-639-318

```

```

Alignment Scores:
Pred. No.: 0 Length: 2157
Score: 631.00 Matches: 717
Percent Similarity: 99.72% Conservative: 1
Best Local Similarity: 99.72% Mismatches: 1
Query Match: 87.88% Indels: 2
DB: 18 Gaps: 0

```

US-10-028-384-6 (1-718) x US-10-793-639-318 (1-2157)

```

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
Db 1 ATGGGATCCGACCGGCTGCTGTTTGTCTGTGTTTTCAGACCATCTCTCAAGCTCGTCATC 60
QY 21 PheValAlaIlePheGlyValAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
Db 61 TTCGTGGCGATTTTGGGGCTGCCATATCATCAAGTTTGTTCAGTCATCAATTTGAG 120
QY 41 SerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrLysTyrLeuValAsn 60
Db 121 TCTATTATCCATGAATTCGACCCCTGTTCAATTATAGGGCTACCAATATATCTCGTCAAC 180
QY 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspAspArgThrTyrTyrProLeuGlyArg 80
Db 181 AATTGGTTTACAAGTTTGTGAACCTGTTTGACGACCTGTTACCTGTTACCTGCGAAGG 240
QY 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrPheIleAla 100
Db 241 GTTACTGGAGGACATTTATATCTCTGTTTGTATGACGACTAGTGTGTTTCTATCTGCGACGC 300
QY 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
Db 301 CTGGCCCACTGGTTGGGCTTCCCATTTGACATCAGAAACGTTTGTGTGTTTGTGGCCCA 360
QY 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla 140
Db 361 CTATTTCCTGGGCTCACCGCTCGGCGACTTACCAATTTACGAAAGAGATTAAGATGCC 420
QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
Db 421 AGCGTGGGCTTTGGCTGCTGTTTATAGCCATTTGCCGTTTCCCGGTTATATATCTAGATCA 480
QY 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
Db 481 GTGGCGGGGCTCTACGATAATGAGCGCATTTGCCATTTACACTATTATTAATGGTCACCTT 540
QY 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
Db 541 TTTTGGATTTAGGCCCAAAAGACTGGCTCTATCATGACGCAACGTTGCGAGCTTTATTTC 600
QY 201 TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
Db 601 TACTTCTACATGGTGTGCGCTTGGGCTGGATGATGTTTCTATCACCACCTTGTATCTCCAC 660
QY 221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
Db 661 CATGCTTTTTCCTGATTTTTCATGGGCGAGATTTCTGTCCTGTTTCTGCTGCTACACC 720
QY 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
Db 721 ACTTGGTACGCTATTGGAACCTGTTGTCATCCATTCAGATCCCATTTGTGCTGCTTACCT 780
QY 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
Db 781 ATCAGGTCTACAGCACCATGCGCGCATTTGGGTGTTTTCGGTGTGTTTTCAGATTTGCGCC 840
QY 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetVal 300
Db 841 TTCGTGACTTCTGTTGAAGGGCCAAATCAGCACAGCTAAGTTTAAAGTCATCATGATGGTT 900

```

QY	301	SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly	320
DB	901	TCTCTGTTTTGATCTTGGTCCCTTGGTGGCGAGACTTCTGCGCTACACTATATGGGG	960
QY	321	LeuIleAlaProTppThrGlyArgpPheTyrSerLeuTtpAspThrAsnTyrAlaLysIle	340
DB	961	TTGATTCGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCACTACGCCAAGATC	1020
QY	341	HisIleProIleIleAlaSerValSerGluHisGlnProValSerTtpProAlaPhePhe	360
DB	1021	CACATTCCTATPACTTGCCTCCGTTTCCGAACATCAACCCGTTTCGTGGCCGGCTTCTTC	1080
QY	361	PheAspThrHisPheLeuIleTtpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp	380
DB	1081	TTTGATACCCACTTTTGTACTGGCTATTTCGCCGCGGTGTATTCCTACTATTCTCGAC	1140
QY	381	LeuIleAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly	400
DB	1141	TTGAAAGACGACGCTTTTTCATCCGCTTACTTCGTTCTGTGTTTGGTACTTTCGGGT	1200
QY	401	ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla	420
DB	1201	GTTATGGTTGATTTGATTTGCACTTTGACACAGCTCATCTGTGTGTCCGCGCGGTGCA	1260
QY	421	LeuSerIysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle	440
DB	1261	TTGTCGAAGATAITTGACATCTACCTGGATTTCAGACAGATGACCGCAATACGCCATC	1320
QY	441	LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr	460
DB	1321	AAACCTGGGCACACTACCTGGGCAAAATGATTTGTTCCGATCATTCATCTTTATTGTAT	1380
QY	461	LeuPheValPheHisSerThrTtpValThrArgThrAlaTyrSerSerProSerValVal	480
DB	1381	CTTTTCGCTTCCATTCTACTTGGTAAACAAGAACTGCATACTCTCTCTCTCTGTGTT	1440
QY	481	LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr	500
DB	1441	TTGCCATCACAAACCCAGATGGTAATTTGGGTGTATCGACACTTCAGGGAAGCGTAC	1500
QY	501	TyrTtpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTtpTtpAspTyrGly	520
DB	1501	TATTGGTTAAGAAATGAACCTCTGATGAGGACAGTAAGGTTCAGCGCTGGTGGGATTC	1560
QY	521	TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTtpAsnAsn	540
DB	1561	TACCAAAATGGTGGCATGGCAGCAGACAGAACCACTTTAGTCGATACCAACGCTGGAA	1620
QY	541	ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle	560
DB	1621	ACTCATCTCGGCATCGTTGGTAAGGCATGGCTTCCCTGAAGAGAAATCTTACGAAAT	1680
QY	561	LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly	580
DB	1681	CTAAAAGAGCATGATGTCGATTATGCTTGGTTCATCTTTGGTGGTCTAATTTGGGTTT	1740
QY	581	GlyAspAspIleAsnLysPheLeuTtpMetIleArgIleSerGluGlyIleTtpProGlu	600
DB	1741	GGTGATGACATCAACAAATCTTGTGGATGATCAGAATTAGCGAGGAATCTGGCCAGAA	1800
QY	601	GluIleLysGluArgTyr-PheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAl	620
DB	1801	GAGATAAAGAGCGGTGA-TTCTATACCGCAGAGGAGAGAAATACAGAGTAGATGCAAGGC	1860
QY	620	AspGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLe	640
DB	1860	TTTCTGAGACCATGAGGAATCTCGTACTTTACAGATGTCTCTCAAAAGATTTTCCCACA	1920
QY	640	uPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal	660
DB	1920	ATTCAATGGTGGCAAGCCACTGACAGAGTGGCTCAACAAATGATCACACCAATTAGACGT	1980

QY	660	lPrpProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyr	680
DB	1980	CCCAACCAATTAGACTACTTCGACGAAGTTTTTACTTCCGAAACACTGGATGGTTAGAAATATA	2039
QY	680	rGlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrAr	700
DB	2040	TCAATTGAAGAAGCATGATGCCCAAGGTAGAACTTTGAGGACAGTTGGTGAGTTTAACCA	2099
QY	700	gSerSerThrIlysthrArgArgSerIleLysArgProGluLeuGlyLeuArgVal	718
DB	2100	GTCTTCTACGAAACCAAGGTCCTATAAGAGACTGATTAAGCTTGGAGATC	2154

RESULT 4	
US-10-032-585-6323	
Sequence 6323, Application US/10032585	
Publication No. US20030180931A1	
GENERAL INFORMATION:	
APPLICANT: Terry, Roemer D.	
APPLICANT: Bo, Jiang	
APPLICANT: Charles, Boone	
APPLICANT: Howard, Bussey	
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery	
FILE REFERENCE: 10182-005-999	
CURRENT APPLICATION NUMBER: US/10/032,585	
CURRENT FILING DATE: 2001-12-20	
NUMBER OF SEQ ID NOS: 8000	
SOFTWARE: Patent version 3.1	
SEQ ID NO 6323	
LENGTH: 2256	
TYPE: DNA	
ORGANISM: Candida albicans	
US-10-032-585-6323	

Alignment Scores:	
Pred. No.:	7,04e-23
Score:	32.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	4.46%
DB:	15
Length:	2256
Matches:	32
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-028-384-6 (1-718) x US-10-032-585-6323 (1-2256)		
QY	322 lleAlaProTrpThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIleHis	341
DB	991 APTGCTCTCGACAGTAGATTTATTCTTATGGGATACAAATTATGCCAAGATTTCAT	1050
QY	342 lleProIleIleAlaSerValSerGluHisGlnPro	353
DB	1051 ATTCAATTATTGCTTCTGTTCTTGACATCAACCT	1086

RESULT 5	
US-10-128-714-2139	
Sequence 2139, Application US/10128714	
Publication No. US20030119013A1	
GENERAL INFORMATION:	
APPLICANT: Jiang, Bo	
APPLICANT: Hu, Wenqi	
APPLICANT: Tishkoff, Daniel	
APPLICANT: Zamudio, Carlos	
APPLICANT: Eroshkin, Alexey M	
APPLICANT: Lemieux, Sebastien M	
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus	
TITLE OF INVENTION: Methods of Use	
FILE REFERENCE: 10182-018-999	
CURRENT APPLICATION NUMBER: US/10/128,714	
CURRENT FILING DATE: 2002-04-23	
PRIOR APPLICATION NUMBER: US 60/285,697	
PRIOR FILING DATE: 2001-04-23	
PRIOR APPLICATION NUMBER: US 60/287,066	
PRIOR FILING DATE: 2001-04-27	
PRIOR APPLICATION NUMBER: US 60/295,890	
PRIOR FILING DATE: 2001-06-05	

Thu Dec 16 16:25:08 2004

; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 2139
 ; LENGTH: 1848
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus
 US-10-128-714-2139

Alignment Scores:
 Pred. No.: 2,18e-11 Length: 1848
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 15 Gaps: 0

US-10-028-384-6 (1-718) x US-10-128-714-2139 (1-1848)

QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 Db 448 TACATGGTGTGGCATGGGTGGGTATGTCCTCATACGAACCTGATCCCTGCAGTT 507
 QY 223 Phe 223
 Db 508 TTT 510

RESULT 6
 US-10-128-714-1139
 ; Sequence 1139, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Wengdi
 ; APPLICANT: Tishkoff, Daniel
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Broshkin, Alexey M
 ; APPLICANT: Lemieux, Sebastien M
 ; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
 ; TITLE OF INVENTION: Methods of Use
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR APPLICATION NUMBER: US 60/287,066
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 1139
 ; LENGTH: 1969
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus

Alignment Scores:
 Pred. No.: 2,32e-11 Length: 1969
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 15 Gaps: 0

US-10-028-384-6 (1-718) x US-10-128-714-1139 (1-1969)
 QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 Db 625 TACATGGTGTGGCATGGGTGGGTATGTCCTCATACGAACCTGATCCCTGCAGTT 684
 QY 223 Phe 223
 Db 685 TTT 687

QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 Db 514 TACATGGTGTGGCATGGGTGGGTATGTCCTCATACGAACCTGATCCCTGCAGTT 573
 QY 223 Phe 223
 Db 574 TTT 576

RESULT 7
 US-10-128-714-7139
 ; Sequence 7139, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Wengdi
 ; APPLICANT: Tishkoff, Daniel
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Broshkin, Alexey M
 ; APPLICANT: Lemieux, Sebastien M
 ; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
 ; TITLE OF INVENTION: Methods of Use
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR APPLICATION NUMBER: US 60/287,066
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 7139
 ; LENGTH: 2232
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus

US-10-128-714-7139
 Alignment Scores:
 Pred. No.: 2,61e-11 Length: 2232
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 15 Gaps: 0

US-10-028-384-6 (1-718) x US-10-128-714-7139 (1-2232)

QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 Db 625 TACATGGTGTGGCATGGGTGGGTATGTCCTCATACGAACCTGATCCCTGCAGTT 684
 QY 223 Phe 223
 Db 685 TTT 687

RESULT 8
 US-10-320-797-2305
 ; Sequence 2305, Application US/10320797
 ; Publication No. US20040014955A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Broshkin, Alexey M.
 ; APPLICANT: Zamudio, Carlos
 ; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOPORMANS AND
 ; TITLE OF INVENTION: METHODS OF USE
 ; FILE REFERENCE: 10182-021-999
 ; CURRENT APPLICATION NUMBER: US/10/320,797
 ; CURRENT FILING DATE: 2002-12-16

; PRIOR APPLICATION NUMBER: 60/341,261
 ; PRIOR FILING DATE: 2001-12-17
 ; NUMBER OF SEQ ID NOS: 3361
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2305
 ; LENGTH: 2466
 ; TYPE: DNA
 ; ORGANISM: Cryptococcus neoformans
 US-10-320-797-2305

Alignment Scores:
 Pred. No.: 2,87e-11 Length: 2466
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 16 Gaps: 0

US-10-028-384-6 (1-718) x US-10-320-797-2305 (1-2466)

QY 152 ileValProGlyTyrIleSerArgSerValalacGlySerTyrAspAsnGluAla 171
 Db 637 ATTGTACCGGATACATCTCGATCTGTCGCGGTTCTTATGACACGAAGCAATGCC 696

QY 172 ile 172
 Db 697 ATC 699

RESULT 9

US-10-128-714-6139
 ; Sequence 6139, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Wengqi
 ; APPLICANT: Tishkoff, Daniel
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Eroshkin, Alexey M
 ; APPLICANT: Lemieux, Sebastien M
 ; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES IN ASPERGILLUS FUMIGATUS AND
 ; TITLE OF INVENTION: METHODS OF USE
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US/10128,714
 ; CURRENT FILING DATE: 2002-04-23
 ; PRIOR APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR APPLICATION NUMBER: US 60/287,066
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 6139
 ; LENGTH: 2603
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus
 US-10-128-714-6139

Alignment Scores:
 Pred. No.: 3.02e-11 Length: 2603
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 15 Gaps: 0

US-10-028-384-6 (1-718) x US-10-128-714-6139 (1-2603)

QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeulleProLeuHisVal 222

Db 877 TACATGGTGTGGCATGGGTGGGTATGTTTCATTACGAACCTGATCCCTCGACGT 936
 QY 223 Phe 223
 Db 937 TTT 939

RESULT 10

US-10-320-797-1305
 ; Sequence 1305, Application US/10320797
 ; Publication No. US20040014955A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eroshkin, Alexey M.
 ; APPLICANT: Zamudio, Carlos
 ; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
 ; TITLE OF INVENTION: METHODS OF USE
 ; FILE REFERENCE: 10182-021-999
 ; CURRENT APPLICATION NUMBER: US/10/320,797
 ; CURRENT FILING DATE: 2002-12-16
 ; PRIOR APPLICATION NUMBER: 60/341,261
 ; PRIOR FILING DATE: 2001-12-17
 ; NUMBER OF SEQ ID NOS: 3361
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1305
 ; LENGTH: 2882
 ; TYPE: DNA
 ; ORGANISM: Cryptococcus neoformans
 US-10-320-797-1305

Alignment Scores:
 Pred. No.: 3.33e-11 Length: 2882
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 16 Gaps: 0

US-10-028-384-6 (1-718) x US-10-320-797-1305 (1-2882)

QY 152 ileValProGlyTyrIleSerArgSerValalacGlySerTyrAspAsnGluAla 171
 Db 796 ATTGTACCGGATACATCTCTCGATCTGTCGCGGTTCTTATGACACGAAGCAATGCC 855

QY 172 ile 172
 Db 856 ATC 858

RESULT 11

US-10-128-714-139
 ; Sequence 139, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Wengqi
 ; APPLICANT: Tishkoff, Daniel
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Eroshkin, Alexey M
 ; APPLICANT: Lemieux, Sebastien M
 ; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES IN ASPERGILLUS FUMIGATUS AND
 ; TITLE OF INVENTION: METHODS OF USE
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US/10/128,714
 ; CURRENT FILING DATE: 2002-04-23
 ; PRIOR APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR APPLICATION NUMBER: US 60/287,066
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31

; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 139
; LENGTH: 3969
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-139

Alignment Scores:
Pred. No.: 4,51e-11 Length: 3969
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.92% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-6 (1-718) x US-10-128-714-139 (1-3969)

Qy 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
Db 1514 TACATGGTGTGGCATGGGTGGGTATGCTTCATTACGAACTGATCCCGCTGCACGTT 1573

Qy 223 Phe 223
Db 1574 TTT 1576

RESULT 12

US-10-128-714-5139
; Sequence 5139, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5139
; LENGTH: 4603
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5139

Alignment Scores:
Pred. No.: 5.19e-11 Length: 4603
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.92% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-6 (1-718) x US-10-128-714-5139 (1-4603)

Qy 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
Db 1877 TACATGGTGTGGCATGGGTGGGTATGCTTCATTACGAACTGATCCCGCTGCACGTT 1936

Qy 223 Phe 223
Db 1937 TTT 1939

RESULT 13

US-10-320-797-305
; Sequence 305, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroshkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; PRIOR FILING DATE: 2002-12-15
; PRIOR APPLICATION NUMBER: 60/341,261
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 305
; LENGTH: 4738
; TYPE: DNA
; ORGANISM: Cryptococcus neoformans
US-10-320-797-305

Alignment Scores:
Pred. No.: 5.34e-11 Length: 4738
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.92% Indels: 0
DB: 16 Gaps: 0

US-10-028-384-6 (1-718) x US-10-320-797-305 (1-4738)

Qy 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171
Db 1652 ATTGTACCGGATACATCTCTCGATCTGTGCGCGTCTTATGACACGAAGCCATTGCC 1711

Qy 172 Ile 172
Db 1712 ATC 1714

RESULT 14

US-10-028-384-7
; Sequence 7, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, Gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 2417
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AF132552
; DATABASE ENTRY DATE: 1999-04-27
; RELEVANT RESIDUES: (1)..(2417)
US-10-028-384-7

Alignment Scores:
Pred. No.: 3.18e-10 Length: 2417
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 14, 2004, 13:16:58 ; Search time 5558.03 Seconds
(without alignments)
4707.376 Million cell updates/sec

Title: US-10-028-384-6
Perfect score: 718
Sequence: 1 MGSRSCLVSVFQTLKLV.....TRSTKRSIKRPELGLRV 718

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 32822875 seqs, 18219865908 residues
Word size: 1
Total number of hits satisfying chosen parameters: 65644297
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Command line parameters:
-MODE=frames+2n.model -DEV=xlh
-Q/cgn2_1/USPIO_spool/US10028384/runat_14122004_131646_6807/app_query.fasta_1.3740
-DB=EST -QFMT=fastcap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn_1.13960_@runat_14122004_131646_6807 -NCPU=6 -ICPU=3
-NO WMAP -LARGECUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -YGAPEXT=60 -Fgapop=6
-Fgapext=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7
Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hcc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gse1.*
9: gb_gse2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Query		Match		Length		DB		ID		Description	
Result No.	Score	Match	Length	DB	ID	Score	Match	Length	DB	ID	Score	Match	Length	DB	ID
1	47	5.5	981	9	CNS06Y0T	47	5.5	981	9	CNS06Y0T	47	5.5	981	9	AL420435 T3 end of
2	41	5.7	869	9	CNS06DYS	41	5.7	869	9	CNS06DYS	41	5.7	869	9	AL394419 T3 end of
3	41	5.7	1088	9	CNS06CLU	41	5.7	1088	9	CNS06CLU	41	5.7	1088	9	AL392680 T7 end of
4	35	4.9	463	8	BZ301622	35	4.9	463	8	BZ301622	35	4.9	463	8	BZ301622 K01378.p1
5	32	4.5	972	9	CNS07A8F	32	4.5	972	9	CNS07A8F	32	4.5	972	9	AL436261 T7 end of
6	29	4.0	1074	9	CNS06ZDA	29	4.0	1074	9	CNS06ZDA	29	4.0	1074	9	AL422180 T3 end of
7	25	3.5	593	8	BZ295215	25	3.5	593	8	BZ295215	25	3.5	593	8	BZ295215 CG1357.r1
8	24	3.3	950	9	CNS06LIS	24	3.3	950	9	CNS06LIS	24	3.3	950	9	AL403622 T3 end of
9	24	3.3	950	9	CNS06L1W	24	3.3	950	9	CNS06L1W	24	3.3	950	9	AL403626 T3 end of

10	20	2.8	424	7	CO337666	10	20	2.8	424	7	CO337666
11	20	2.8	576	4	BM290810	11	20	2.8	576	4	BM290810
12	20	2.8	585	1	AI133998	12	20	2.8	585	1	AI133998
13	20	2.8	594	4	BM638349	13	20	2.8	594	4	BM638349
14	20	2.8	598	4	BI170396	14	20	2.8	598	4	BI170396
15	20	2.8	600	4	BI163592	15	20	2.8	600	4	BI163592
16	20	2.8	615	4	BI364613	16	20	2.8	615	4	BI364613
17	20	2.8	616	2	BF504378	17	20	2.8	616	2	BF504378
18	20	2.8	630	1	AI295381	18	20	2.8	630	1	AI295381
19	20	2.8	635	1	AI135629	19	20	2.8	635	1	AI135629
20	20	2.8	641	1	AI257750	20	20	2.8	641	1	AI257750
21	20	2.8	642	4	BI173041	21	20	2.8	642	4	BI173041
22	20	2.8	644	4	BG636414	22	20	2.8	644	4	BG636414
23	20	2.8	649	4	BG641064	23	20	2.8	649	4	BG641064
24	20	2.8	652	4	BI171940	24	20	2.8	652	4	BI171940
25	20	2.8	654	4	BI374334	25	20	2.8	654	4	BI374334
26	20	2.8	655	2	BF496296	26	20	2.8	655	2	BF496296
27	20	2.8	655	4	BI484774	27	20	2.8	655	4	BI484774
28	20	2.8	658	4	BI374189	28	20	2.8	658	4	BI374189
29	20	2.8	663	4	BG641172	29	20	2.8	663	4	BG641172
30	20	2.8	666	4	BI357074	30	20	2.8	666	4	BI357074
31	20	2.8	668	4	BI227902	31	20	2.8	668	4	BI227902
32	20	2.8	679	7	CK657649	32	20	2.8	679	7	CK657649
33	20	2.8	681	7	CK659033	33	20	2.8	681	7	CK659033
34	20	2.8	682	7	CK659064	34	20	2.8	682	7	CK659064
35	20	2.8	700	4	BM595496	35	20	2.8	700	4	BM595496
36	20	2.8	715	7	CK657451	36	20	2.8	715	7	CK657451
37	20	2.8	718	6	CD779819	37	20	2.8	718	6	CD779819
38	20	2.8	736	1	AA949890	38	20	2.8	736	1	AA949890
39	20	2.8	784	6	CD783059	39	20	2.8	784	6	CD783059
40	20	2.8	788	6	CD781399	40	20	2.8	788	6	CD781399
41	20	2.8	810	2	BF502026	41	20	2.8	810	2	BF502026
42	20	2.8	876	6	CD794446	42	20	2.8	876	6	CD794446
43	20	2.8	878	6	CD788595	43	20	2.8	878	6	CD788595
44	20	2.8	891	6	CD778215	44	20	2.8	891	6	CD778215
45	20	2.8	960	6	CD793779	45	20	2.8	960	6	CD793779

ALIGNMENTS

RESULT 1
CNS06Y0T
LOCUS
DEFINITION
T3 end of clone AY0005C11 of library AY00A from strain CBS 6340
of Kluyveromyces thermotolerans, genomic survey sequence.
AL420435
AL420435.1 GI:12203620
GSS
Kluyveromyces thermotolerans
Kluyveromyces thermotolerans
Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Kluyveromyces.
REFERENCE
1 (bases 1 to 981)
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Neuvéglise, C., Olier-Kalogeropoulos, O., Portier, S.,
Saurin, W., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEMS Lett. 487 (1), 3-12 (2000)
JOURNAL
MEDLINE
PUBMED
20584711
11152876
REFERENCE
2 (bases 1 to 981)
Malpertuy, A., Llorente, B., Blandin, G., Artiguenave, F., Wincker, P.
and Dujon, B.
Genomic exploration of the hemiascomycetous yeasts: 10.
Kluyveromyces thermotolerans
FEMS Lett. 487 (1), 61-65 (2000)
JOURNAL
MEDLINE
PUBMED
20584720
11152885
REFERENCE
3 (bases 1 to 981)

AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES Location/Qualifiers
 1..981
 /organism="Kluyveromyces thermotolerans"
 /mol_type="genomic DNA"
 /strain="CBS 6340"
 /db_xref="taxon:4916"
 /clone="AY0AA005C11"
 /clone_lib="AY0AA"
 /note="end : T3"
 misc_feature
 <7..5939
 /note="similar to Saccharomyces cerevisiae ORF YGL022w [
 STT3 ; oligosaccharyl transferase subunit]"
 /evidence=not_experimental

ORIGIN
 Alignment Scores:
 Pred. No.: 4e-38 Length: 981
 Score: 47.00 Matches: 47
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.55% Indels: 0
 DB: 9 Gaps: 0

US-10-028-384-6 (1-718) x CNS06Y0T (1-981)
 QY 365 PheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAspGlu 384
 Db 88 TTTCGATTGGTATTCCCGGGGTATTTTACGTTCTCTGCTCTTATTTTCGAGTCAAGATGA 147
 QY 385 HisValPheValIleAlaTrpSerValLeuCySserTyrPheAlaGlyValMetValarg 404
 Db 148 CACGTTTTCGTGATTGCTTACTCCGTTCTCTGCTCTTATTTTCGCGGTGTCATGTCAGA 207
 QY 405 LeuMetLeuThrLeuThrPro 411
 Db 208 CTCATGCTAACTCTGACACCA 228

RESULT 2
 CNS06DY5/c
 LOCUS 869 bp DNA linear GSS 17-JUN-2001
 DEFINITION T3 end of clone AR0AA017A04 of library AR0AA from strain CBS 732 of
 Zygosaccharomyces rouxii, genomic survey sequence.
 ACCESSION AL394419
 VERSION AL394419.1 GI:12145378
 KEYWORDS GSS.
 ORGANISM Zygosaccharomyces rouxii
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.
 1 (bases 1 to 869)
 Souciat,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorante,B.,
 Malpertuy,A., Neuvéglise,C., Olier-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 869)
AUTHORS de Montigny,J., Straub,M., Potier,S., Tekala,F., Dujon,B.,
 Wincker,P., Artiguenave,F. and Souciat,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 8.
 Zygosaccharomyces rouxii
 FEBS Lett. 487 (1), 52-55 (2000)
JOURNAL 20584718
MEDLINE 11152883
PUBMED 11152883
REFERENCE 3 (bases 1 to 869)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES Location/Qualifiers
 1..869
 /organism="Zygosaccharomyces rouxii"
 /mol_type="genomic DNA"
 /strain="CBS 732"
 /db_xref="taxon:4956"
 /clone="AR0AA017A04"
 /clone_lib="AR0AA"
 /note="end : T3"
 misc_feature
 complement(<2..5775)
 /note="similar to Saccharomyces cerevisiae ORF YGL022w [
 STT3 ; oligosaccharyl transferase subunit]
 2 putative frameshift(s)"
 /evidence=not_experimental

ORIGIN
 Alignment Scores:
 Pred. No.: 7.51e-32 Length: 869
 Score: 41.00 Matches: 41
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.71% Indels: 0
 DB: 9 Gaps: 0

US-10-028-384-6 (1-718) x CNS06DY5 (1-869)
 QY 151 AlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIle 170
 Db 335 GCGATTGGCCAGGTACATTTCTAGATCAGTTCAGGTTCTTACGATACGAGCTATT 276
 QY 171 AlaIleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnIleThrGlySer 190
 Db 275 GCAATTACCCCTTTTAATGGTTACTTTTCATGTTCTGGATTAAAGCTCAAGACAGGTCT 216
 QY 191 Ile 191
 Db 215 ATT 213

RESULT 3
 CNS06CLU/c
 LOCUS 1088 bp DNA linear GSS 14-JUN-2001
 DEFINITION T7 end of clone AR0AA006C05 of library AR0AA from strain CBS 732 of
 Zygosaccharomyces rouxii, genomic survey sequence.
 ACCESSION AL392680
 VERSION AL392680.1 GI:12142260
 KEYWORDS GSS.

```

SOURCE      Zygosaccharomyces rouxii
ORGANISM    Zygosaccharomyces rouxii
REFERENCE   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS     Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.
1 (bases 1 to 1088)
Souciet, J.-L., Aglie, M., Artiguenave, F., Blandin, G.,
Bollotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrien, P., Lepingle, A., Lorente, B.,
Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekai, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
11152876
REFERENCE   2 (bases 1 to 1088)
AUTHORS     de Montigny, J., Straub, M., Potier, S., Tekai, F., Dujon, B.,
Wincker, P., Artiguenave, F. and Souciet, J.
Genomic exploration of the hemiascomycetous yeasts: 8.
Zygosaccharomyces rouxii
FEBS Lett. 487 (1), 52-55 (2000)
20584718
11152883
REFERENCE   3 (bases 1 to 1088)
AUTHORS     Genoscope.
Direct Submission
Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
secref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomycetes bayanus var. uvarum, Saccharomycetes
exiguus, Saccharomycetes servazzii, Zygosaccharomycetes rouxii,
Saccharomycetes kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
    source
        1..1088
            /organism="Zygosaccharomyces rouxii"
            /mol_type="genomic DNA"
            /strain="CBS 732"
            /db_xref="taxon:4956"
            /clone="AR0A006C05"
            /clone_lib="AR0AA"
            /note="end : 17"
misc_feature
    complement( <3..>1006)
        /note="Similar to Saccharomycetes cerevisiae ORF YGL022w [
        SRT3 ; oligosaccharyl transferase subunit ]
        1 putative frameshift (8)
        /evidence=not_experimental

ALIGNMENT Scores:
Pred. No.:      9,39e-32      Length:      1088
Score:          41.00        Matches:      41
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     5.71%        Indels:      0
DB:              9           Gaps:        0

US-10-028-384-6 (1-718) x CNS06CLU (1-1088)

QY 151 AlalievAlprCGlyTyrIleSerArgSerValalaglySerTyAspAsnGluAlalle 170
Db 566 GCGATTGTGCCAGGTTACATTCCTAGATCAGTTCAGGTTCTTACGATAACGAGCTATT 507

QY 171 AlalievAlprCGlyTyrIleSerArgSerValalaglySerTyAspAsnGluAlalle 190
Db 506 GCAATTACCCCTBTAAATGGTTTACTTTTCATGTTCTCGGATTAAGCTCAAAAGACAGGTTCT 447

SOURCE      Zygosaccharomyces delphensis
ORGANISM    Zygosaccharomyces delphensis
REFERENCE   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS     Kluyveromyces delphensis
1 (bases 1 to 463)
Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.
Evidence from comparative genomics for a complete sexual cycle in
the 'asexual' pathogenic yeast Candida glabrata
Genome Biol. 4 (2), R10 (2003)
22508158
12620120
COMMENT     Contact: Wong S
            Department of Genetics, Smurfit Institute
            Trinity College Dublin
            Dublin 2, Ireland
            Tel: 353 1 6082319
            Fax: 353 1 6798558
            Email: swong@tcd.ie
            Class: plasmid ends.
FEATURES
    source
        1..463
            /organism="Kluyveromyces delphensis"
            /mol_type="genomic DNA"
            /strain="CBS 2170"
            /db_xref="taxon:51657"
            /clone="KD1378"
            /clone_lib="Kluyveromyces delphensis Random Genomic
            Library"

ALIGNMENT Scores:
Pred. No.:      8.49e-26      Length:      463
Score:          35.00        Matches:      35
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     4.87%        Indels:      0
DB:              8           Gaps:        0

US-10-028-384-6 (1-718) x BZ301622 (1-463)

QY 320 GlyLeuileAlaprotipThrGlyArgPheTySerLeuTrpaspThrAsnTyAlaLys 339
Db 160 GGTTCGATCGCGCCATCGACTCGAAGATTTTACTCTTTGGGATACCAATACGCTAAG 219

QY 340 IleHisIleProIleIleAlaSerValSerGluHisGluProVal 354
Db 220 ATTCAATCCCAATCAATGGCTGTGCTCTGGAACATCAACCTGTT 264

RESULT 5
CNS07A8F
LOCUS
DEFINITION T7 end of clone BC0AA002D03 of library BC0AA from strain CBS 767 of
            Debaryomyces hansenii, genomic survey sequence.
ACCESSION AL436261
VERSION   AL436261.1
KEYWORDS  GI:12219674
SOURCE    Debaryomyces hansenii (anamorph: Candida famata)
ORGANISM  Debaryomyces hansenii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

```

Saccharomycetales; Saccharomycetaceae; Debaryomyces.
1 (bases 1 to 972)
LOCUS      Saccharomycetaceae; Debaryomyces.
DEFINITION
AUTHORS    Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
JOURNAL    FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE    20584711
PUBMED     1152876
REFERENCE  2 (bases 1 to 972)
AUTHORS    Lepingle,A., Casaregola,S., Neuveglise,C., Bon,E., Nguyen,H.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 14.
            Debaryomyces hansenii var. hansenii
JOURNAL    FEBS Lett. 487 (1), 82-86 (2000)
MEDLINE    20584724
PUBMED     1152889
REFERENCE  3 (bases 1 to 972)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT    This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES   Location/Qualifiers
            source          1..972
                        /organism="Debaryomyces hansenii"
                        /mol_type="genomic DNA"
                        /strain="CBS 767"
                        /variety="hansenii"
                        /db_xref="taxon:4959"
                        /clone="BC0AA002D03"
                        /clone_lib="BC02A"
                        /note="end : T7"
            misc_feature    <5..>970
                        /note="similar to Saccharomyces cerevisiae ORF YGL022w [
                        STT3 ; oligosaccharyl transferase subunit ]"
                        /evidence=not_experimental

ORIGIN
Alignment Scores:
Pred. No.:      2,59e-22      Length:      972
Score:          32.00%      Matches:      32
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    4.4e%      Indels:      0
DB:              9      Gaps:        0

US-10-028-384-6 (1-718) x CNS07A8F (1-972)

QY      322   llaalaprtrprrghlyavghpethyrsrleutipasthrasntrhalaystlehis 341
Db       143   attgcacttggatgcgcctttctatctttatggatgactaaactatcccaagattcac 202
QY      342   ileprollelelaalservalsergluHisglnpro 353
Db       203   attcctataattgcattccgittcttgaacatcaacct 238

RESULT 6
CNS06ZDA

```

```

CNS06ZDA      1074 bp      DNA      linear      GSS 06-JUL-2001
T3 end of clone XAY0AA001F10 of library XAY0AA from strain CBS 6340
of Kluyveromyces thermotolerans, genomic survey sequence.
ACCESSION    AL422180
VERSION      AL422180.1 GI:12205374
KEYWORDS     GSS.
SOURCE       Kluyveromyces thermotolerans
ORGANISM     Kluyveromyces thermotolerans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE    1 (bases 1 to 1074)
AUTHORS      Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
TITLE        Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
JOURNAL      FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE      20584711
PUBMED       1152876
REFERENCE    2 (bases 1 to 1074)
AUTHORS      Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.
            and Dujon,B.
TITLE        Genomic exploration of the hemiascomycetous yeasts: 10.
            Kluyveromyces thermotolerans
JOURNAL      FEBS Lett. 487 (1), 61-65 (2000)
MEDLINE      20584720
PUBMED       1152885
REFERENCE    3 (bases 1 to 1074)
AUTHORS      Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (09-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT      This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES     Location/Qualifiers
            source          1..1074
                        /organism="Kluyveromyces thermotolerans"
                        /mol_type="genomic DNA"
                        /strain="CBS 6340"
                        /db_xref="taxon:4916"
                        /clone="XAY0AA001F10"
                        /clone_lib="XAY0AA"
                        /note="end : T3"
            misc_feature    <882..>1066
                        /note="similar to Saccharomyces cerevisiae ORF YGL022w [
                        STT3 ; oligosaccharyl transferase subunit ]"
                        /evidence=not_experimental

ORIGIN
Alignment Scores:
Pred. No.:      4,16e-19      Length:      1074
Score:          29.00%      Matches:      29
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    4.04%      Indels:      0
DB:              9      Gaps:        0

US-10-028-384-6 (1-718) x CNS06ZDA (1-1074)

QY      63   phtyrllyspheleuasnrtrpheaaspargthrtrptyrproleuglyargvalthr 82
CNS06ZDA

```



```

Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 3.34%             Indels: 0
DB: 9                           Gaps: 0

US-10-028-384-6 (1-718) x CNS06L1W (1-950)

QY 578 G7PhedClyAAspaSsileasntysPheUmrPmEileargileSerGlucGlylle 597
DB 77 GGTTTCGGTGGTGATGATATTAATAAAGTTTTTATGGATGATCAAGATTAGTGAAGTTAT 136

QY 598 TrpProGluGlu 601
DB 137 TGGCCAGAAAGA 148

RESULT 10
C0337666 424 bp mRNA linear EST 29-JUN-2004
LOCUS EN15033:Sprime Exelixis FlyTag MN08 Bluescript Drosophila
DEFINITION melanogaster cDNA clone EN15033 5, mRNA sequence.
ACCESSION C0337666
VERSION C0337666.1 GI:49397941
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 424)
AUTHORS Nakanishi,M., Muzong,C., Peterson,E., Laufer,A., Leung,W., Platt,D.,
and Swimmer,C.
TITLE Exelixis FlyTag EST Project MN08 Library
JOURNAL Unpublished (2004)
COMMENT Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: EN150 row: C column: 9
High quality sequence stop: 243.
Location/Qualifiers
1..424
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="EN15033"
/cell_line="mbn2"
/clone_lib="Exelixis FlyTag MN08 Bluescript"
/notes="Vector: pBluescript, Site 1: NotI; Site 2: XhoI;
oligoat primed from LSP induced mbn2 cell line."

ORIGIN
Alignment Scores:
Pred. No.: 5,07e-10 Length: 424
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.79% Indels: 0
DB: 7 Gaps: 0

US-10-028-384-6 (1-718) x C0337666 (1-424)

QY 149 PheileAlaIleValProGlyTyrileSerArgSerValAlaGlySerTyrAspAsnGlu 168
DB 56 TTCATCGCCATCGTGCCTGGCTACATCAGTAGTTCGGTGGCTGCATACGATACGAG 115

RESULT 11
C03376810 576 bp mRNA linear EST 01-JUL-2002
LOCUS BM290810
DEFINITION BM290810 AVSG Amblyomma variegatum cDNA clone AVAAR06 5' end, mRNA
sequence.
ACCESSION BM290810

```

```

VERSION BM290810.1 GI:21640777
KEYWORDS EST.
SOURCE Amblyomma variegatum
ORGANISM Amblyomma variegatum

REFERENCE
AUTHORS Nene,V., Lee,D., Quackenbush,J., Skilton,R., Mwaura,S.,
Gardner,M.J., and Bishop,R.
TITLE AVGI, an index of genes transcribed in the salivary glands of the
ixodid tick Amblyomma variegatum
JOURNAL Int. J. Parasitol. 32 (12), 1447-1456 (2002)
MEDLINE 22281296
PUBMED 12392910
COMMENT Contact: Vish Nene
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-610-5968
Fax: 301-838-0208
Email: nene@tigr.org
Seq primer: M13 reverse.
FEATURES
source
1..576
Location/Qualifiers
/organism="Amblyomma variegatum"
/mol_type="mRNA"
/db_xref="taxon:34610"
/clone="AVAR06"
/tissue_type="Salivary glands"
/dev_stage="Adult"
/lab_host="E.coli strain DH10B-Tona"
/clone_lib="AvSG"
/notes="Vector: PCMV-SPORT6.1; Salivary glands were
dissected on day five after initiation of feeding. Total
RNA was prepared using acid guanidium
thiocyanate-phenol-chloroform extraction. The cDNA library
was custom prepared by Invitrogen Corporation. Briefly,
first strand cDNA was primed using oligo(dT) containing a
NotI site. Size fractionated double stranded cDNA was
ligated to EcoRV-NotI cleaved vector and electroporated
into E.coli."
ORIGIN
Alignment Scores:
Pred. No.: 6,886-10 Length: 576
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.79% Indels: 0
DB: 4 Gaps: 0
US-10-028-384-6 (1-718) x BM290810 (1-576)

RESULT 12
AI133998
LOCUS AI133998 585 bp mRNA linear EST 19-APR-2001
DEFINITION GH1327.5prime GH Drosophila melanogaster head pOT2 Drosophila
melanogaster cDNA clone GH1327 5prime, mRNA sequence.
ACCESSION AI133998
VERSION AI133998.1 GI:3626556
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Phytozoa; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 585)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.

TITLE BGP/HIMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit P element sequence 1(3)j2D9
Plate: 113 row: C column: 3
High quality sequence stop: 470.
FEATURES
source
1..585
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="GH1327"
/sex="male and female"
/dev_stage="adult"
/lab_host="PH5 - alpha"
/clone_lib="GH Drosophila melanogaster head pOT2"
/notes="Organ: head; Vector: pOT2; Site 1: EcoRI; Site 2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2. Plasmid cDNA library."
ORIGIN
Alignment Scores:
Pred. No.: 6,996-10 Length: 585
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.79% Indels: 0
DB: 1 Gaps: 0
US-10-028-384-6 (1-718) x AI133998 (1-585)

QY 149 PheilealalevalProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168
|||||
Db 506 TTCATGCCATTCGCTGGCTACATCAGTAGTCGGTGGATCGTACGATACGAG 565
|||||

RESULT 13
BM638349
LOCUS BM638349 594 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687566421 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
BM638349
ACCESSION BM638349
VERSION BM638349.1 GI:18937860
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE 1 (bases 1 to 594)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Citera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Citera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltR@citera.com
Plate: NU01003CY8 row: 0 column: 23
Seq primer: M13 Reverse.
FEATURES
source
1..594
Location/Qualifiers
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"

```

ORIGIN

/clone="19600449630257"
 /dev_stage="Adult"
 /lab_host="DH10B"
 /clone_lib="A.Gam.ad.cDNA1"
 /note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen.
 cDNA inserts >500 bp cloned directionally into pSport 1.
 Not 1 site is 3'. Clones available through the Malaria
 Research and Reference, Reagent Resource Center
 (www.malaria.mr4.org)."

Alignment Scores:

Pred. No.:	7,1e-10	Length:	594
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.79%	Indels:	0
DB:	4	Gaps:	0

US-10-028-384-6 (1-718) x BM638349 (1-594)

Qy 149 PheilleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168

Db 320 TTCATCGGATCGTCCCGGCTACATCAGCCGCTCGTGGCGGCTCGTACGATACAGAG 379

RESULT 14

Bi170396

LOCUS

DEFINITION

RE11825.Sprime RE Drosophila melanogaster normalized Embryo pF1c-1
 Drosophila melanogaster cDNA clone RE11825 5 similar to OstStt3:
 FBan007748 'enzyme' located on: 3R 96B16-96B17;; 04/12/2001, mRNA
 sequence.

ACCESSION

Bi170396

VERSION

Bi170396.1 GI:14636203

KEYWORDS

EST.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 598)

AUTHORS

Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
 Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
 George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
 Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
 Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
 Rubin,G.M.

TITLE

BDGP/HMI RE Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 hit genomic AE003750: arm:3R [20671983,20899380]
 estimated-cyto:96B10-96C1: 04/12/2001 hit P element 1(3)j2D9.
 1(3)j2D9 AQ026308 inserted at base 292 5' end of P element Inverse
 PCR: 03/18/2001
 Plate: RE.118 row: C column: 1
 High quality sequence stop: 547.
 Location/Qualifiers

FEATURES

source

1..598

/organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE11825"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DH5-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pF1c-1"

/note="Organ: embryo; Vector: pF1c1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Alignment Scores:

Pred. No.:	7,1e-10	Length:	598
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.79%	Indels:	0
DB:	4	Gaps:	0

US-10-028-384-6 (1-718) x Bi170396 (1-598)

Qy 149 PheilleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168

Db 483 TTCATCGGATCGTCCCGGCTACATCAGTGGTGGTGGCTGATACGATACAGAG 542

RESULT 15

Bi163592

LOCUS

DEFINITION

RE03026.Sprime RE Drosophila melanogaster normalized Embryo pF1c-1
 Drosophila melanogaster cDNA clone RE03026 5 similar to OstStt3:
 FBan007748 'enzyme' located on: 3R 96B16-96B17;; 04/11/2001, mRNA
 sequence.

ACCESSION

Bi163592

VERSION

Bi163592.1 GI:14629399

KEYWORDS

EST.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 600)

AUTHORS

Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
 Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
 George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
 Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
 Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
 Rubin,G.M.

TITLE

BDGP/HMI RE Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 hit genomic AE003750: arm:3R [20671983,20899380]
 estimated-cyto:96B10-96C1: 04/11/2001 hit P element 1(3)j2D9.
 1(3)j2D9 AQ026308 inserted at base 292 5' end of P element Inverse
 PCR: 03/17/2001
 Plate: RE.30 row: C column: 2
 High quality sequence stop: 455.
 Location/Qualifiers

FEATURES

source

1..600

/organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE03026"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DH5-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pF1c-1"

/note="Organ: embryo; Vector: pF1c1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Alignment Scores:

Pred. No.:	7.17e-10	Length:	600
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.79%	Indels:	0
DB:	4	Gaps:	0

US-10-028-384-6 (1-718) X B1163592 (1-600)

Qy 149 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168

Db 499 TTCATCGCCATCGTCCCTGGCTACATCACTAGTGGTGGCTGGATCGTACGATACGAG 558

Search completed: December 15, 2004, 06:14:10
Job time : 5570.03 secs

1292

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:51:45 ; Search time 7954.29 Seconds
(without alignments)
4268.640 Million cell updates/sec

Title: US-10-028-384-6
Perfect score: 3761
Sequence: 1 MGSDFSCVLSVFOTILKLV.....TRSSSTKRSIKRPELGLRV 718

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DSV=xlh
-O=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131519_8017/app.query.fasta_1.3740
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=ptt -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptt -NORV=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn_1.1.20262@runat_14122004_131519_8017 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -YGAPEXT=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3761	100.0	2733	6	AX799086 Sequence
2	3761	100.0	2733	8	D28952 Saccharomyc
3	3751	99.7	3633	8	SCYGL022W
4	3488	92.7	2000	6	AX595714 Sequence

5	3488	92.7	2000	6	AX819898	AX819898 Sequence
6	3488	92.7	2000	6	AX830928	AX830928 Sequence
7	2969	78.9	110000	8	CR380947_0	CR380947 Candida g
8	2905	77.2	110000	8	CR382123_09	Continuation (11 o
9	2905	77.2	110000	8	CR382123_10	Continuation (11 o
10	2901	77.1	300029	8	AE016903	AE016903 Eramothec
11	2738	72.8	2256	6	AX489023	AX489023 Sequence
12	2708.5	72.0	110000	8	CR382134_03	Continuation (4 of
13	2495.5	66.4	110000	8	CR382128_21	Continuation (22 o
14	2237	59.5	40907	8	SPBC1271	AL034353 S.pombe c
15	2224.5	59.1	5048	8	AB015232	AB015232 Schizosac
16	2180.5	58.0	94169	8	BX842635	Neurospor
17	2027.5	53.9	2417	3	AF132552	AF132552 Drosophil
18	2027.5	53.9	2417	3	AX799088	AX799088 Sequence
19	2027.5	53.9	2699	6	CQ589353	CQ589353 Sequence
20	2016.5	53.6	4236	10	BC052433	BC052433 Mus muscu
21	2013.5	53.5	2710	6	AX799084	AX799084 Sequence
22	2012.5	53.5	2481	6	AX799082	AX799082 Sequence
23	2012.5	53.5	2481	9	AY074880	AY074880 Homo sapi
24	1966	52.0	2567	5	BC063234	BC063234 Dario rer
25	1956	52.0	2608	5	BC046072	BC046072 Dario rer
26	1955.5	52.0	4922	6	CQ589352	CQ589352 Sequence
27	1955.5	52.0	162921	3	AC007853	AC007853 Drosophil
28	1955.5	52.0	181132	3	AC008206	AC008206 Drosophil
29	1955.5	52.0	227219	3	AE003750	AE003750 Drosophil
30	1949	51.8	2118	12	BT008132	BT008132 Synthetic
31	1946	51.7	2118	9	BT007100	BT007100 Homo sapi
32	1946	51.7	2450	9	BC048348	BC048348 Homo sapi
33	1946	51.7	2458	6	CQ834714	CQ834714 Sequence
34	1946	51.7	2472	6	CQ727867	CQ727867 Sequence
35	1946	51.7	2516	9	BC020965	BC020965 Homo sapi
36	1945	51.7	2293	9	HUMTWC	L47337 Homo sapien
37	1945	51.7	2736	10	BC037612	BC037612 Mus muscu
38	1945	51.7	3094	6	AX799090	AX799090 Sequence
39	1945	51.7	3094	10	MUSSTRPR	L34260 Mus musculu
40	1940.5	51.6	2481	5	BC067313	BC067313 xenopus t
41	1937	51.5	3666	10	AK129027	AK129027 Mus muscu
42	1934	51.4	2898	9	HSMB09254	BS649102 Homo sapi
43	1928.5	51.3	2435	6	CQ834710	CQ834710 Sequence
44	1928.5	51.3	75650	2	AC018145	AC018145 Drosophil
45	1920	51.1	2472	6	AX799092	AX799092 Sequence

ALIGNMENTS

RESULT 1
AX799086
LOCUS AX799086 2733 bp DNA linear PAT 08-OCT-2003
DEFINITION Sequence 5 from Patent WO03054008.
ACCESSION AX799086
VERSION AX799086.1 GI:37605060
KEYWORDS
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE
AUTHORS Perreault, C. and McBride, K.
TITLE Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy
JOURNAL Patent; WO 03054008-A 5 03-JUL-2003;
COMPATIGENE Inc. (CA)
FEATURES
Location/Qualifiers
1..2733
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/note="STT3, Acc# D28952"

ORIGIN

Alignment Scores:
Pred. No.: 7.58e-305 Length: 2733
Score: 3761.00 Matches: 718

Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-028-384-6 (1-718) x AX799086 (1-2733)

QY	1	MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLeuLeuValIle	20
DB	219	ATGGGATCCGACCGGTGCGTGTGTTTGTCTGTGTTTTCAGACCACTCTCAAGCTGTCATC	278
QY	21	PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLeuPheGlu	40
DB	279	TTCTGTCGGCATTTTGGGCTGCCATATCATCACCTTTGTTGTCAGTCATCAAAATTTGAG	338
QY	41	SerIleIleHisGluPheAspProPheAsnTyrArgAlaThrLysTyrLeuValAsn	60
DB	339	TCTATTATCCATGAATTCGACCCCTGGTTCAATATATAGGGCTACCAAAATATCTCGTCAC	398
QY	61	AsnSerPheTyrLysPheLeuAsnTrpPheAspAspArgThrTrpTyrProLeuGlyArg	80
DB	399	AATTCGTTTTTACAAGTTTTTGAAGTTTGTGACGACCGTACCTGGTACCCCTCGGAAGG	458
QY	81	ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTrpHisAla	100
DB	459	GTTACTGGAGGACTTTATATCTCTGGTTTGTATGACGACTAGTGGCTTCATCTGGCAGGCC	518
QY	101	LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro	120
DB	519	CTGGCGCACTGGTGGCTTGGCCATTGCATCAGAAACGTTGTGTCTATTGGCGCA	578
QY	121	LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla	140
DB	579	CTATTTTCTGGGTCAACCCCTGGCGACTTACGAATTTACGAAAGAGATTAAAGATGCC	638
QY	141	SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer	160
DB	639	AGCGCTGGGCTTTGGCTGCTGGTTTATAGCCATTGTCGGGTATATATCTAGATCA	698
QY	161	ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet	180
DB	699	GTGGCGGGTCTCTACGATTAATGAGCCATTGGCCATTACACTATTAAATGCTCATTATG	758
QY	181	PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe	200
DB	759	TTTTGGATTAAAGCCCAAAAGACTGGCTCTATCATGCGCAACGTGTCAGCTTTATTC	818
QY	201	TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeu	220
DB	819	TACTTCTACATGCTGTCGGCTTGGGTGGATACGTGTTTCATCACCAACTTGATCCCACTC	878
QY	221	HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr	240
DB	879	CATGCTCTTTTGTGATTGATGGGACAGATATTCGTCCAACTGATTCTGCCCTACACC	938
QY	241	ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro	260
DB	939	ACTTGTGACGTATTGGAACCTGTTCATCCATCCAGATCCCATTTTGGTTTCTCTACTC	998
QY	261	IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla	280
DB	999	ATCAGTCTTAACACACATGCGCGCATTTGGGTGTTTTCGGTTTTCATTTCAGATGTCGCC	1058
QY	281	PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal	300
DB	1059	TTCCGGTGACTTCGTGAAGGGCCCAATCACACAGCTAGTTTAAAGTCATCATGATGTT	1118
QY	301	SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly	320
DB	1119	TCTCTCTTTTGTATCTGGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1178
QY	321	LeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle	340

DB	1179	TTGATTGCCCTTGGACTGGTAGATTTTATTCTGTTATGGGATACCACATAGCCAAAGATC	1238
QY	341	HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProAlaPhePhe	360
DB	1239	CACATTCTTATCATGCTCCGTTTCCGAAACATCAACCGGTTTCGTGGCCGCTTCTTC	1298
QY	361	PheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp	380
DB	1299	TTTGATACCCACTTTTGTATCTGGCTATTTCCTCCGCGGTGTATTCTCTACTATTCTCGAC	1358
QY	381	LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly	400
DB	1359	TTGAAAGACGACGACGTTTTTGTCTATCGCTTACTCCGTTCTGTGTCTGTTCTGCCGT	1418
QY	401	ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla	420
DB	1419	GTTATGTTTAGATTGATGTTGACTTTGACACCACTCATCTGTGTGTCGCGCCGTCGCA	1478
QY	421	LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle	440
DB	1479	TTGTCGAAGATATTGACATCTACCTGGATTTCAGACCAAGTACCCGCAATACGCCATC	1538
QY	441	LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr	460
DB	1539	AAACCTGGCGCACTACTGSCCAATTTGATGTTTTCGGATCATTCATCTTTATTGTTAT	1598
QY	461	LeuPheValPheHisSerThrTrpValThrAlaTyrSerSerProSerValVal	480
DB	1599	CTTTTCGCTCTTCCATTCTACTTGGGTAAACAGAACTGCATCTCTCTCTCTCTCTCT	1658
QY	481	LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr	500
DB	1659	TTGCCATCACAAACCCAGATGTTAAATGGCGTTGATGCGACACTTCAGGGAACGTTAC	1718
QY	501	TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaIleTrpAspTyrGly	520
DB	1719	TATTGTTTGAAGATGAACCTCATGAGGACAGTAGGTTCAGCGTGTGGGATTACCGT	1778
QY	521	TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn	540
DB	1779	TACCAAAATTTGGTGGCATGCGACAGACCACTTTAGTCGATTAACAACACGCTGGAACAT	1838
QY	541	ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle	560
DB	1839	ACTCATCATCGCCATCGTTGGTAAAGCCATGGCTTCCCTGGAAGAGAAATCTTACGAAAT	1898
QY	561	LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly	580
DB	1899	CTAAAAGACGATGATGCGATTATGCTTTGGTCTATCTTTGGTGGTCTAATTCGGGTTGT	1958
QY	581	GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu	600
DB	1959	GGTGTATGATCATCAACAATTTCTTGGATGATCAGNAATTAGCGAGGAATCTGGCCAGAA	2018
QY	601	GluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAla	620
DB	2019	GAGATAAAAGAGCGTTTATTTCTATACCGCAGAGGAGGAATACAGAGTAGATCAAGGGCT	2078
QY	621	SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu	640
DB	2079	TCGAGACCATGAGGAACCTCGCTACTTTTACAGATGTCCTACAAAGATTTCCCAATTA	2138
QY	641	PheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAspVal	660
DB	2139	TTCAATGGTGGCCCAAGCCACTGACAGAGTGGCTCAACAAATGATCACCATTAGACGTC	2198
QY	661	ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyr	680
DB	2199	CCACCATTAGACACTTCGACGAGTTTTTACTTCCGAAATTCGATGTTGATGATATAT	2258
QY	681	GlnLeuLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyLeuLeuThrArg	700
DB	2259	CAATTGAAGAGGATGATGCCAAGGTAGAACTTTGAGGAGCGTTGTTGAGTTAACGAG	2318

1253	Db	ATGGGATCCGACCGGTGGTGTGTTGTCGTGTGTTTCAGACCATCCTCAAGCTCGTCATC	1312
21	QY	PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu	40
1313	Db	TTCTGTGGCGAATTTTGGGGTGGCATATCATCACGTTTGTGTTCAGTCATCAAAATTTGAG	1372
41	QY	SerIleIleHisGluPheAspProTTPheAsnTyrArgAlaThrLysTyrLeuValAsn	60
1373	Db	TCTATTATCCATGAATTCGACCCCTGGTTCATTTATAGGGCTPACCAAAATCTCGTCAC	1432
61	QY	AsnSerPheTyrLysPheLeuAsnTTPheAspArgThrTTPyrProLeuGlyArg	80
1433	Db	TAATTCGTTTACAAGTTTTTGAATCGTTTGACGACCGGTACCTGTGTACCCCTCGGAAGG	1492
81	QY	ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTTPHisAla	100
1493	Db	GTTACTGGAGGACCTTTATATCTGTTTGTATGAAGCATAGTGGCTTCACTCGCACGCC	1552
101	QY	LeuArgAsnTTPLeuGlyLeuProIleAspIleA:GAsnValCysValLeuPheAlaPro	120
1553	Db	CTGGCCAACTGGTTGGGCTTGGCCATGTGACATCAGAAACGTTTGTGTCTATTTCGGCCA	1612
121	QY	LeuPheSerGlyValThrAlaTTPAlaThrTyrGluPheThrLysGluIleLysAspAla	140
1613	Db	CTATTTCTGGGCTCACCGCTGGGCGACCTTACGAATTTACGAAGAGATTTAAGATGCC	1672
141	QY	SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer	160
1673	Db	AGCGTGGGCTTTTGGGCTGTGTGTTTATAGCAATGTCCCGGTATATATCTPAGATCA	1732
161	QY	ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet	180
1733	Db	GTTGGCGGGTCTCTACGATATATGAGGCATTTGCCATTTCACTATTATATGGTCATTTCTG	1792
181	QY	PheTTPLeLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe	200
1793	Db	TTTTTGGATTAAAGGCCCAAGACCTGGCTCTATCATGCACGCAACGTGTGCAGCTTTATTTC	1852
201	QY	TyrPheTyrMetValSerAlaTTPGlyGlyTyrValPheIleThrAsnLeuIleProLeu	220
1853	Db	TACTTCTACATGGTGTGGCTTGGGGTGAGTACGTGTTCATCCCACTTGATCCCACTC	1912
221	QY	HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr	240
1913	Db	CATGTCTTTTTCGTGATTTTTCATGGCGCAGATATTGTCCAAACTGTATTCTGCCTACCC	1972
241	QY	ThrTTPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro	260
1973	Db	ACTTGGTACGCTATTGAACTGTTCATCCATGCAGATCCCAATTGTTCGGTTTCTTACCT	2032
261	QY	IleArgSerAsnAspHisMetAlaLeuLeuGlyValPheGlyLeuIleGlnIleValAla	280
2033	Db	ATCAGGTCTACGACCAACCATGGCGCATTTGGGTGTTTTCGGTTTGATTTCAGATTGTCCGC	2092
281	QY	PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal	300
2093	Db	TTCCGTGTACTTCGTGMAAGGCCAAATCAGCAGACGTAAAGTTTAAAGTCATCATGTGGTT	2152
301	QY	SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly	320
2153	Db	TCTCTGTTTTCGTATCTGGTCCCTTGGTGTGGTCGACCTTCTCGCTTGACCTATATGGGG	2212
321	QY	LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTTPAspThrAsnTyrAlaLysIle	340
2213	Db	TTGATTCGCCCTTGGACCTGGTAGATTTTATCGTTATGGGATACCACTACGCAAGATC	2272
341	QY	HisIleProIleIleAlaSerValSerGluHisGlnProValSerTTPProAlaPhePhe	360
2273	Db	CACATTCCTATCATTTGCTCCGTTTCCGAACATCAACCCGTTTTCGTGGCCGCTTTCTTC	2332
361	QY	PheAspThrHisPheLeuIleTTPLeuPheProAlaGlyValPheLeuLeuPheLeuAsp	380

RESULT 4
AX595714
LOCUS

AX595714	2000 bp	DNA	linear	PAT 14-FEB-2003
----------	---------	-----	--------	-----------------


```

Db      961  TTGATGGCCCTTGACCTGGTAGATTTTATTGGTATGGGATACCACTACGCAAGATC 1020
QY      341  HisIleProIleAlaSerValSerGluHisGlnProValSerTyrProAlaPhePhe 360
Db      1021  CACATTCCTATCATTCGCTCGTTTCCGAACATCAACCGCTTTCGTGGCCGCTTCTTC 1080
QY      361  PheAspThrHisPheLeuIleThrLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
Db      1081  TTGTATACCATCTTTTGTATCGGTATTCCTCCGCGGTGATTCCTACATATTCCTGCAG 1140
QY      381  LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
Db      1141  TTGAAGACGACGACGCTTTTGTGTCATCGTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1200
QY      401  ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValala 420
Db      1201  GTTATGGTTAGATTGATGTGATCTTGACACGAGTCATCTGTGTGTCGCGCGCGTCGCA 1260
QY      421  LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
Db      1261  TTGTCCACGATATTTGACATCTACCTGGATTTCAAGACAGTACCGCCAAATACGCCATC 1320
QY      441  LysProAlaAlaLeuAlaLysIleValSerGlySerPheIlePheTyrLeuTyr 460
Db      1321  AAACCTGCGGCACTACTGCGCAAAATGATGTGTTCCGGATCATTCATCTTTTATTGTAT 1380
QY      461  LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerValVal 480
Db      1381  CTTTTCGCTCTCCATCTACTTGGTAAAGACAGTGCATCTCTCTCTCTCTCTCTCTCT 1440
QY      481  LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
Db      1441  TTGCCATCAAAACCCAGATGGTAAATTTGGCTTTGATGACGACACTTCAGGAGCGGTAC 1500
QY      501  TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTyrTrpAspTyrGly 520
Db      1501  TATTGGTTAAGAAATGAATCTGATGAGACAGTAAAGTTGACGCGTGTGGGATTTACGGT 1560
QY      521  TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
Db      1561  TACCAATTTGGTGGATGCGACAGAACCACTTTAGTCGATACACACAGCTGGAAACAT 1620
QY      541  ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle 560
Db      1621  ACTCATCGCCATCGTTGGTAAAGCCATGCTCCCTGAGAGAAATCTTACGAAAT 1680
QY      561  LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
Db      1681  CTAAAGACGATGATGCGATATGCTTTGGTCACTTTTGTGTCTCAATTTGGTTGGT 1740
QY      581  GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTyrProGlu 600
Db      1741  GGTGATGACATCAACAAATTTCTTGTGGATGATCAGAAATAGCAGGGAATCTGCCAGAA 1800
QY      601  GluIleLysGluArgTyrPheTyrThrIleGluGlyGluTyrArgValAlaPheAlaArgAla 620
Db      1801  GAGATAAAGACGCTGATTTCTATACCGCAGGAGGAATACAGAGTAGATGCAAGGGCT 1860
QY      621  SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu 640
Db      1861  TCTGAGACCATGAGGAACCTCGTACTTTTACAAGATGCTCTACAAGATTTCCCAACATTA 1920
QY      641  PheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAspVal 660
Db      1921  TTCAATGGTGGCCAGCCCACTGACAGAGTGGTCAACAAATGATCACACCATTAGACGTC 1980
QY      661  ProProLeuAspTyrPhe 666
Db      1981  CCACCATTTAGACTACTTC 1998

```

RESULT 7
CR380947_0

WPCOMMENT
Sequence split into 5 fragments LOCUS CR380947 Accession CR380947
Fragment Name Begin End
CR380947_0 1 110000
CR380947_1 100001 210000
CR380947_2 200001 310000
CR380947_3 300001 410000
CR380947_4 400001 485192

LOCUS CR380947 485192 bp DNA linear PLN 30-JUL-2004
DEFINITION Candida glabrata strain CBS138 chromosome A complete sequence.
ACCESSION CR380947
VERSION
KEYWORDS
SOURCE
ORGANISM
Candida glabrata CBS138
Candida glabrata CBS138
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
AUTHORS
1 Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Blevaksten, C., Boisarane, A., Boyer, J., Cattolico, L., Confantolieri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantreya, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potter, S., Richard, G.F., Straub, M.L., Suleau, A., Swennen, D., Tekala, F., Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P. and Souclet, J.L.
Genolevures
Genome evolution in yeasts
Nature 430 (6995), 35-44 (2004)
15229592

2
Frangeul, L. and Sherman, D.
Direct Submission
Submitted (05-MAY-2004) B. Dujon, Institut Pasteur, Unite de
Genetique Moleculaire des Levures, 25 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: lfrangeul@pasteur.fr E-mail:
david@labri.fr
This is the first release of the complete genome sequence. The
Candida glabrata sequence is made of 13 chromosomes (A to M). These
sequences contains 6 gaps (artificially closed) indicated by 'gap'
features and about ten colinearity problems indicated by 'gap'
'misc difference' features. An update of the sequence will be
submitted in few weeks. For more information go to
http://cbi.labri.fr/Genolevures.
Location/Qualifiers
1. 485192
/organism="Candida glabrata CBS138"
/mol_type="genomic DNA"
/strain="CBS138"
/db_xref="taxon:284593"
/chromosome="A"
complement(1608..2765)
/locus_tag="CAGLOA00110g"
/pseudo
complement(1608..2765)
/locus_tag="CAGLOA00110g"
/note="some similarities with sp|P08640 Saccharomyces
cerevisiae YK019c STA1, no start, frameshift"
/pseudo
complement(2671..4809)
/locus_tag="CAGLOA00121g"
/note="some similarities with sp|P38894 Saccharomyces

CONSRMT
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
JOURNAL
COMMENT
FEATURES
source
gene
CDS
gene
CDS

QY 43 IleHisGluPheAspProTyrPheAsnTyrArgAlaThrLysTyrLeuValAsnAsnSer 62
Db 22392 ATTATGATATGACCCCTGGTTCAACTTCAGGCTACAAATAACCTGGTTTCACATCG 22451
QY 63 PheTyrLysPheLeuAsnTyrPheAspAspArgThrTyrPyrProLeuGlyValThr 82
Db 22452 TTCACGAGTTCCTGAACTGGTCGACGATAGACATGGTATCAATGGGTGAGTCACT 22511
QY 83 GlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAlaLeuArg 102
Db 22512 GGTGTACTTGTACCCCTGGTTGATGACCAAGTGTCTTTGATGGCAGTCTGGCGC 22571
QY 103 AsnTyrLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaProLeuPhe 122
Db 22572 AAG---ATCGGTCTACCTGTTGACATTAGAAACATCTGTGCTTTTGTCTGCTTTC 22628
QY 123 SerGlyValThrAlaThrAlaThrTyrGluPheThrLysGluIleLysAspAlaSerAla 142
Db 22629 GCGGTATTACCGTATCGCACTTACCAATTCGAAGGAATGAAGGATTCCTCGGC 22688
QY 143 GlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSerValAla 162
Db 22689 GGTCTACTGGCTGCTGCTCATGCTATTCTCCAGGTTACATTTCCAGATCTGTGCT 22748
QY 163 GlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMetPheTyr 182
Db 22749 GGTCTTACGATACGAAGCATCTCAATTAATCTATGATGAATCAATTCATGTTCTGG 22808
QY 183 IleLysAlaGlnLysThrGlySerIleMethHisAlaThrCysAlaAlaLeuPheTyrPhe 202
Db 22809 ATCAAAGCACAAAGACTGGGTCCATCATGCAAGCCACTTTCGCTGCTTTTCTACTTT 22868
QY 203 TyrMetValSerAlaThrGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
Db 22869 TACATGGTCTCAGCTTGGGGTGGTACGCTCTTACTTACTTAAATTCATACACGTT 22928
QY 223 PheLeuLeuIleLeuMetGlyArgTyrSerLysLeuTyrSerAlaTyrThrTyrTyr 242
Db 22929 TTTATCCTTATCCTAATGGCGGTTATAACTCTAGATTATACATGCTTACTCCACTGG 22988
QY 243 TyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleArg 262
Db 22989 TACGCAATGGTACATGGCTTCATGCAAAATCCITTTGGGGTTTCTTCCCAATCAGA 23048
QY 263 SerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGly 282
Db 23049 TCTATGACCATATGGCTGCTCGGTGTTTCGGCTTCAATCBAATGTTGCTTTGGT 23108
QY 283 AspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeu 302
Db 23109 GACTTCGTCAGGCGCAGAGTCCCCGCACAAAGTCAAGGTCTATCTCCTTGGTTTCTTTA 23168
QY 303 PheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIle 322
Db 23169 TTGGTGTGATGGCTGTAGGTATTTTGGGTTTGTGCTTACTTACTTCCGTTTGAAT 23228
QY 323 AlaProTyrThrGlyArgPheTyrSerLeuTyrAspThrAsnTyrAlaLysIleHisIle 342
Db 23229 GCGCATGACCTGGTGGATTTTACTCTCTTGGGATACGCTACGTAAGGTTCACT 23288
QY 343 ProIleIleAlaSerValSerGluHisGlnProValSerTyrProAlaPhePheAsp 362
Db 23289 CCTATCAATGCTTCTGTCGGAACATCAACAGTAGCTTGGCATCTTCTTTTGTGAC 23348
QY 363 ThrHisPheLeuIleThrLeuPheProAlaGlyValPheLeuLeuPheLeuLys 382
Db 23349 CAACACTTTTGTATGCTGTGTTCCAGCGGTGTTTCTTGTGTTTCATGGAATGAAG 23408
QY 383 AspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGlyValMet 402
Db 23409 GATGACGACGCTTGTATCTGTATTCGCTACTCTGTTTCATCTACTGCTGTGTATG 23468
QY 403 ValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSer 422

Db 23469 ATTAGATGATGTTGACCTTAACTCCAATTTGTTGTCTCTCTGCTGCTATGCTTTATCA 23528
QY 423 LysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLys---TyrAlaIle--- 440
Db 23529 AAAATATTGATATCTACTTAGATTCTCAAGAGAGATGCTGATGGTAAAGCGGTCCA 23588
QY 441 -----LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePhe 457
Db 23589 AGTTTGGACAGCCAGCATCCATTTGGCCAGTTGGTATCTCTGTTTCTCTTTATCTTC 23648
QY 458 TyrLeuTyrLeuPheValPheHisSerThrTyrValThrArgThrAlaTyrSerSerPro 477
Db 23649 TATCTATACCTGTACGCTACACTGTACCTGGGTCACTAAGACAGCTACTCTCTCTCT 23708
QY 478 SerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspPheArg 497
Db 23709 TCTGTGTTCTTGAACACAGCAGACCAAGTTGGTAACAATATTAATCACTGATCTTAGA 23768
QY 498 GluAlaTyrTyrTyrLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTyrPyr 517
Db 23769 GAAGCCTATTACTGGCTAAGAAATCAACACTGCTGAAGACGCTAAGATTGCTCTGCTGG 23828
QY 518 AspTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnThr 537
Db 23829 GATTACGTTTACCAAAATCGGTGGTCTTGCAGACAGACCACTTGTGTTGACAAATAC 23889
QY 538 TrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGlyLysSer 557
Db 23889 TGAACAATACACATATCGCATTCGTTAAAGCAATGCTGCTCCAGAAAGATTCT 23948
QY 558 TyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyLeuIle 577
Db 23949 TATAGTTATTGAAGAACAATGATTTGATGCTCTGCTGCTCTTGTGTTGTTT 24008
QY 578 GlyPheGlyGlyAspAspIleAsnLysPheLeuTyrMetIleArgIleSerGluGlyLe 597
Db 24009 GGTTCGTTGCTGCTGACGATATCAATAGTTCCTATGGATGATTAGATCTCTAGTGTATC 24068
QY 598 TrpProGluGlyLysGluArgTyrPheTyrThrAlaGluGlyLeuTyrArgValAsp 617
Db 24069 TGGCCAGATGATGATCCAAAGAAAGAACTATTTCAAGTACCTGAGTACAGATGTGAC 24128
QY 618 AlaArgAlaSerGluThrMetArgAsnSerLeuLysTyrLysMetSerTyrLysAspPhe 637
Db 24129 GCCATGGCTCTAGACCATGAGCAATCTGTGATGTACAAGATGTGCTACCAAGGCTTT 24188
QY 638 ProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrPro 657
Db 24189 GCAGACTTGTTCCTACGGCGAGCGCTCGATAGGTTTCGTAAACCAACAGATCGCAAC 24248
QY 658 LeuAspValProLeuAspTyrPheAspGluValPheThrSerGluAsnTyrMetVal 677
Db 24249 TCTCAGCTGAGCCATAGATTATTCGATGAAGTCTTCACTTCCGAGAACTGGTTAGTC 24308
QY 678 ArgIleTyrGlnLeuLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGlu 697
Db 24309 AGATTATACAGCTTAAAGGATGAAGATCCAGAAAGGTAGAACTACGCGACGCTCGGATTG 24369
QY 698 LeuThrArg---SerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeu 716
Db 24369 ATCAACAGACATCCAGCTCATTTCAGAAAGAACTCTCTGAAGAAACCAAGAACTCAACATG 24428
QY 717 ArgVal 718
Db 24429 AGAATA 24434

RESULT 8
CR382123_09/c
WPCOMMENT
Sequence split into 18 fragments LOCUS CR382123 Accession CR382123
Fragment Name Begin End
CR382123_00 1 110000

CR382123_01 100001 210000
 CR382123_02 200001 310000
 CR382123_03 300001 410000
 CR382123_04 400001 510000
 CR382123_05 500001 610000
 CR382123_06 600001 710000
 CR382123_07 700001 810000
 CR382123_08 800001 910000
 CR382123_09 900001 1010000
 CR382123_10 1000001 1110000
 CR382123_11 1100001 1210000
 CR382123_12 1200001 1310000
 CR382123_13 1300001 1410000
 CR382123_14 1400001 1510000
 CR382123_15 1500001 1610000
 CR382123_16 1600001 1710000
 CR382123_17 1700001 1753957
 Continuation (10 of 18) of CR382123 from base 900001 (CR382123, Kluyveromyces lactis str)

Alignment Scores:
 Pred. No.: 4,31e-231 Length: 110000
 Score: 2905.00 Matches: 546
 Percent Similarity: 83.16% Conservatives: 81
 Best Local Similarity: 72.41% Mismatches: 79
 Query Match: 77.24% Indels: 48
 DB: 8 Gaps: 6

US-10-028-384-6 (1-718) x CR382123_09 (1-110000)
 QY 9 LeuSerValPheGlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAla 28
 DB 105230 TTGGAGGGTTCCAAACAGPACTCAAGGCCAATCTTTATTGCTATCTTTGGTGGCA 105171
 QY 29 IleSerArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspPro 48
 DB 105170 ATCTCTCAGCTTGTCTCAGTGATTTAGATTTAGCTTATTCACGAGTTCGATCCA 105111
 QY 49 TrpPheAsnTyrArgAlaThrLysTyrLeuValAsnSerPheTyrLysPheLeuAsn 68
 DB 105110 TGGTTTAAATTCAGAGGCCAACAAATATCTAGTCAGTCATTCATTTATGAGTTTTGAAC 105051
 QY 69 TrpPheAspArgThrTyrProLeuGlyArgValThrGlyThrLeuTyrPro 88
 DB 105050 TGGTTTACGATAGGACATGGTATCCCTCGGAGAGTTACCGGTGGTACTTTGTATCCT 104991
 QY 89 GlyLeuMetThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuPro 108
 DB 104990 GGTTTGATGACAACTTCTGGTGCAATTTGGCATATATCTTCTGTAAG---ATTGGTTGGCA 104934
 QY 109 IleAspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrp 128
 DB 104933 ATTGATATTCGTACATTTGTGTGTGTCGACCTGCAATCTCAGGATTCACCGCTGG 104874
 QY 129 AlaThrTyrGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuAlaAlaGly 148
 DB 104873 GCCACGTACGAGTTTACTAAAGAAATTGAAAGATTCTAGTGCAGGTTATTAGCAGCAGCT 104814
 QY 149 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168
 DB 104813 TTTAAGCAATTTGCCCTTGGGTACATTTCTAGATCTGTGGCCGGTTCTTACGACAAATGAA 104754
 QY 169 AlaIleAlaIleThrLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThr 188
 DB 104753 GCCAATGCCATCACCCTTACTAATGTTAAGCTTTATGTTCTGGATCAAAAGCTCAGAAGACA 104694
 QY 189 GlySerIleMetHisAlaThrCysAlaAlaLeuPheThrPheTyrMetValSerAlaTrp 208
 DB 104693 GGTCTATATGTACTCCATGTTAGCGCTTTGTCTTACTTCTACTAGATGGTATCAGCCGCG 104634
 QY 209 GlyGlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMet 228
 DB 104633 GGTGGTTACGTATTTATTACAAACTTAATACCACTACATGTTTTCATTTTGTATCTTGATG 104574

QY 229 GlyArgTyrSerSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrVal 248
 DB 104573 GGACGCTACCAATCGAAACTATATTCGTTATACATGAGTGGACCAATGGTACTTTTA 104514
 QY 249 AlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAla 268
 DB 104513 GCTTCAATGCAAAATTCCTTCGTGGGATTTTGGCAATTAGATTAACCATCATCATGGCT 104454
 QY 269 AlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGln 288
 DB 104453 GCATTGGGTGTTTTTGGTTTAAATCCAAATTTGCACTAGTAGTACTACATCAATCTCAA 104394
 QY 289 IleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuIleLeuValLeu 308
 DB 104393 GTTTCAGGTGAAGATTCAAGGTTGTTATGTTTCAATTTAGGATTTGATTTCTTCTAT 104334
 QY 309 GlyValValGlyLeuSerAlaLeuThrTyrMetClyLeuIleAlaProTyrThrGlyArg 328
 DB 104333 GGTATTTGAGGCTTCTCTTTTAACTTACATGGTTACATTCACCATGGACAGGTAGA 104274
 QY 329 PheTyrSerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerVal 348
 DB 104273 TTCTACTCTCTATGGACACATAATTATGCCAAATTCATATACCTATCATCTGATCTGTG 104214
 QY 349 SerGluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrp 368
 DB 104213 TCTGAACACACAGCTCTCTGCAAGGCCCTCTTTCTTTGATTAACCAATTTTGTATTTGG 104154
 QY 369 LeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheVal 388
 DB 104153 TTGTTCCAGCTGGTGTCTTCTGTTCTCTAGAACATAAGGATGAACATGATGATTCGTA 104094
 QY 389 IleAlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThr 408
 DB 104093 CTGGCTTACTCTGTTTATGCTCATCTTCTGCTGGTGTGATGATCAATTAATGTTGACA 104034
 QY 409 LeuThrProValIleCysValSerAlaAlaValAlaLeuSerLysLysIlePheAspIleTyr 428
 DB 104033 TTGACTCCGATATCTGTGCTACCTGGTGGATGCAATTTCTAAACATTTGATGATTTAT 103974
 QY 429 LeuAspPhe-----LysThrSerAspArgLysTyrAla----- 431
 DB 103973 TTGGATTTCTGTGCGCACTTTTACCAATCTCCCAAGACGTGAAGTCAGTTAACTC 103914
 QY 432-----LysThrSerAspArgLysTyrAla----- 439
 DB 103913 GATAAGGAGAAAGAAAGCAACATCTGATTTAGAACACAGAGATGATGACTATGAAGCTGAA 103854
 QY 440-----IleLysProAlaAla-----LeuLeuAlaLysLeu 449
 DB 103853 TTGGTTTGGATTATGAACCCCTTCTTAAGCGCGTTGGTTGACATCTGCTAAGGCC 103794
 QY 450 IleValSerGlySerPheIlePheTyrIleuPheValPheHisSerThrTrpVal 469
 DB 103793 GTTGTATCATCCACTTTTGTATGATCTATCTCTTTTATCTGCTGCTGCTGCTGCTGCTG 103734
 QY 470 ThrArgThrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyLys 489
 DB 103733 ACATCCAACTTCTACTTCTCCCACTCTGTGGTCATGCACTCAAAATCCAGACGATCG 103674
 QY 490 LeuAlaLeuIleAspAspPheArgGluAlaTyrTrpLeuArgMetAsnSerAspGlu 509
 DB 103673 GTTGTCTTAAATGACGACTTTAGAGAAGCATATTTACTGGTACGTATGATGATGATGAT 103614
 QY 510 AspSerLysValAlaAlaIleTrpAspTyrGlyTyrGlnIleGlyGlyMetAlaAspArg 529
 DB 103613 GATGCCAAGTTGCGCTTGTGGATTTATGTTATCAATCGGTGGTATGGCTGATAGA 103554
 QY 530 ThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAla 549
 DB 103553 ACCACTCTAGTTGATTAATAACACTTGAATAACCAACCACTTGTGTTGTTGTAAGGCA 103494
 QY 550 MetAlaSerProGluLysSerTyrGlyIleLeuLysGluHisAspValAspTyrVal 569

```

Db 103493 ATGTCCTCTCCAGAGAGAAAGCTTACGAATTTTGAAGGCTCATGAGCTTGATATGTC 103434
QY 570 LeuValillePheGlyLeuGlyPheGlyGlyAspAspIleAsnLysPheLeuTyr 589
Db 103433 TTGGTTGTCTCGGTGCTCTAGGATATAGTGTGACGATCTTAATAAATCTTATGG 103374
QY 590 MetIleArgIleSerGluGlyIleTyrProGluGluIleLysGluArgTyrPheTyrThr 609
Db 103373 ATGGTCAGAAATCTCTGAGGCAATTTGGCCAGATGAGTGAAGAGAGCGTAATTTCTTCA 103314
QY 610 AlaGluGlyGluTyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeu 629
Db 103313 GGCAGAGGTGAATACCGTGTTCATGCCAGCGCCCAAGACAAATCAAAAGATAGCTTAATG 103254
QY 630 TyrLysMetSerTyrLysAspPheProGluGluPheAsnGlyGlyGlnAlaThrAspArg 649
Db 103253 TACAAGCTATCTGACCAATTTCCATCTATATATATCAAAATGGTGTAGTCATGATAGA 103194
QY 650 ValArgGlnGluMetIleThrProLeuAspValProProLeuAspTyrPheAspGluVal 669
Db 103193 ACCCGTAACCAAAAGATCACTAACCAATGAAGTTGGCCCTCTTGACTACTTTGAAGAAGTA 103134
QY 670 PheThrSerGluAsnTyrMetValArgIleTyrGlnLeuLysLysAspAlaGlnGly 689
Db 103133 TTCACCTTCGAATCTTGTGTGTCAGAAATTTACAGCTGAATCTCCAGATTCCTTAGGC 103074
QY 690 ArgThrLeuArgAspValGlyGluLeuThrArg-----SerThrLys 704
Db 103073 -----AGAGATCATGGGAGCTGCCAAATTTGACAGAACTTTACAAAGTGGAGTG 103023
QY 705 ThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal 718
Db 103022 AAAAGAGATTAGTGAAGAACCAAGATTTAGACTAGATAGTA 102981

```

RESULT 9

CR382123_10/c

WPCOMMENT

Sequence split into 18 fragments LOCUS CR382123 Accession CR382123

Fragment Name	Begin	End
CR382123_00	1	110000
CR382123_01	100001	210000
CR382123_02	200001	310000
CR382123_03	300001	410000
CR382123_04	400001	510000
CR382123_05	500001	610000
CR382123_06	600001	710000
CR382123_07	700001	810000
CR382123_08	800001	910000
CR382123_09	900001	1010000
CR382123_10	1000001	1110000
CR382123_11	1100001	1210000
CR382123_12	1200001	1310000
CR382123_13	1300001	1410000
CR382123_14	1400001	1510000
CR382123_15	1500001	1610000
CR382123_16	1600001	1710000
CR382123_17	1700001	1753957

Continuation (11 of 18) of CR382123 from base 1000001 (CR382123 Kluveromyces lactis str

Alignment Scores:

Pred. No.:	4,318-231	Length:	110000
Score:	2905.00	Matches:	546
Percent Similarity:	83.16%	Conservative:	81
Best Local Similarity:	72.41%	Mismatches:	79
Query Match:	77.24%	Indels:	48
DB:	8	Gaps:	6

US-10-028-384-6 (1-718) x CR382123_10 (1-110000)

QY 9 LeuSerValPheGlnThrIleLeuLysLeuValillePheValAlaIlePheGlyValAla 28

Db 5230 TTGGAGGTTTCCAAACAGACTACTCAAGGCCCTTAATCTTTATGCTATCTTTGGTGCGCA 5171

```

QY 29 IleSerSerArgLeuPheAlaValilleLysPheGluSerIleIleHisGluPheAspPro 48
Db 5170 ATCTCTTCCAGCTTTGTTCTCAGTGAATTTAGATCTATTATCCACGAGTTTCGATCCA 5111
QY 49 TrpPheAsnTyrArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsn 68
Db 5110 TGGTTTAATTTTCAGAGCCACCNAATATCTAGTCAGTCATTCATTTATGATGTTTTGAAC 5051
QY 59 TrpPheAspAspArgThrTyrProLeuGlyArgValThrGlyGlyThrLeuTyrPro 88
Db 5050 TGGTTTTCAGTAGACATGATATCCCTCGGACAGATTTACCGTGGTACTTTGTATCCT 4991
QY 89 GlyLeuMetThrThrSerAlaPheIleTyrPheIleAlaLeuArgAsnTrpLeuGlyLeuPro 108
Db 4990 GGTITGATGACAACTTCTGGTGCAATTTGGCATATATCTTCGTAAG---ATTGGTTTGCCA 4934
QY 109 IleAspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrp 128
Db 4933 ATTGATATTCGTAACATTTGTGTGTTCTCGACCTGCATTTCTCAGGATTCACCGCTTG 4874
QY 129 AlaThrTyrGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaGly 148
Db 4873 GCCACGTACGAGTTTACTAAAGATTCTAGTCACGGTTATTAGCAGCAGCT 4814
QY 149 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168
Db 4813 TTTATGGCAATGGCCCTCGGTGATCTTCTAGATCTGTGGCCGGTTCTTACGCAATGAA 4754
QY 169 AlaIleAlaIleThrLeuLeuMetValThrPheMetPheTyrIleLysAlaGlnLysThr 188
Db 4753 GCCATTCGCACTACCTTACTTAATGGTAACGTTTATGTTCTCGATCAAGCTCAGAGACA 4694
QY 189 GlySerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrp 208
Db 4693 GGTTCATATGATCTACTCCATGTTAGCAGCTTGTCTTACTTCTACATGGTATCAGCGCTG 4634
QY 209 GlyGlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMet 328
Db 4633 GGTGTTACGTATTTATTACAACTTAATACCACCTACATGTTTTCATTGATCTTGATG 4574
QY 229 GlyArgTyrSerSerLysLeuTyrSerAlaTyrThrTyrTrpTyrAlaIleGlyThrVal 248
Db 4573 GGACGCTACCAATCGAAACTATATTCGTCTTACTACATGGTACCAATTTGGTACTTTA 4514
QY 249 AlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAla 268
Db 4513 GCTTCAATGCAAAATTCCTTTCGTGGGATTTTGCCTAATTAGATCTTAACGATCAGCTGCT 4454
QY 269 AlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGln 288
Db 4453 GCATGGGTGTTTGTGTTTAAATCCAAATTTGTCAGTGGTACATCAAAATCTCAA 4394
QY 289 IleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeu 308
Db 4393 GTTTCAGGTGAAGATTCAGGTGTTGTTATGTTTCATTTCATTAGGATTTGATTTCTATT 4334
QY 309 GlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProThrPheGlyArg 328
Db 4333 GGTATTTTCAGGCTTGTTCTTTTAACTTACATGGGTTACATGGGTTACATGGGACAGATGA 4274
QY 329 PheTyrSerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerVal 348
Db 4273 TTCTACTCTCTATGGACACTAATATGCCAAATTCATATACCTATCATTTGCTATGCTG 4214
QY 349 SerGluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrp 368
Db 4213 TCTGAACACCAAGCCTCTCTGATGGCCCTCTTTTCTTTGATAACCAATTTTGTGATTTGG 4154
QY 369 LeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheVal 388
Db 4153 TTGTTCCAGCTGGTGCTTCTTGTGTTCTTCTAGAACTAAAGGATCAACATGATTCGTA 4094

```


Qy 26 GlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGluSerIleIleHisGlu 45
Db 112532 GAGGGGCAATCTCGTACGACTGTTTCGGTAATCCGGTTCGAGTCCATCATCCACGAG 112591
Qy 46 PheAspProTrpPheAsnTyrArgAlaThrIleLysTyrLeuValAsnAsnSerPheTyrIys 65
Db 112592 TTCGACCCCTGGTGTAACTTCGCGCAACACAGTACCTGGTGGAGCCACTCGTTCACGAG 112651
Qy 66 PheLeuAsnTrpPheAspAspArgThrTrpTyrProLeuGlyArgValThrGlyGlyThr 85
Db 112652 TTCTCGAATTGCTTTCAGCAGCGGAGCTGGTACCGCTAGACCGCTGAGCGGGGGAGC 112711
Qy 86 LeuTyrProGlyLeuMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeu 105
Db 112712 CTGTACCCGGGCTCATACACACTTCTGCGTGTATCTGGCACCGGTTCGGAAGTTC--- 112768
Qy 106 GlyLeuProIleAspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyVal 125
Db 112769 GGGCTGCCAGTGGACATCCGGAACATCTGTGTCTGTTTCGCGCCCGTTCGCGGGCTG 112828
Qy 126 ThrAlaTrpAlaThrTyrGluPheThrIleLysAlaSerAlaGlyLeuLeu 145
Db 112829 ACGGCTGTGGACGATAGAGTTCACGAGGAATCAAGGACGAGAGCGGGCTGTG 112888
Qy 146 AlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyr 165
Db 112889 GCTCGGGCTTCATCGCGATAGTCCCGGTACATCTCGCGGTCTGTGGCGGGCTCTTAC 112948
Qy 166 AspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMetPheTrpIleLysAla 185
Db 112949 GACACAGGAGGATCGCCATCACGCTGCTGATGTCTACGTTCATGTTTCGATCAAGCC 113008
Qy 186 GlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetVal 205
Db 113009 ATCAAAACGGGCTCGATACCTCACTCCACGCTAGCGCGCTCTTCTACTTCTACATG 113068
Qy 206 SerAlaTrpGlyGlyValPheIleThrAsnLeuIleProLeuHisValPheLeuLeu 225
Db 113069 TCTGCTGGGAGGCTAGCTTTCATACGAACTGATCCGCTGCGGTTCATCTCTG 113128
Qy 226 IleLeuMetGlyArgTyrSerSerIleLysTyrSerAlaTyrThrTrpTyrAlaIle 245
Db 113129 ATTCTGATGGGCGGTCAAGAGTAAGCTGTACAGCGGTACACTACGTGGTATCGATC 113188
Qy 246 GlyThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAsp 265
Db 113189 GGGACTGTGGCTCCATCCAGATACCGTTCGTTGGGTTCGCGATTTCGCTCCATGAC 113248
Qy 266 HisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheVal 285
Db 113249 CACATGGCTGGCTTCGGGCTTCGCGCTTCATCCAGTGTGCTGGGTGGGACACTACATG 113308
Qy 286 LysGlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuIle 305
Db 113309 AAGTCCCATGCTCCACCCCAAGTCAAGATCCACTCCATCTCTCCGCGCGCTGTGT 113368
Qy 306 LeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrp 325
Db 113369 GGGGGGATGGAGTGTGGGCTGTGTGCGCTGACCATGTGCGGTTCATCGCTCTTGG 113428
Qy 326 ThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIle 345
Db 113429 ACCGGCGGTTCCTGCTTGGGACACAACTACGGAAGATCCACATCCCATTCAT 113488
Qy 346 AlaSerValSerGluHisGlnProValSerTrpProAlaPhePheAspThrHisPhe 365
Db 113489 GCCTCGGTGTCCGAACACACACCTACGCGGTGGCTGCTCATTTCTTTCACATCAGTIT 113548
Qy 366 LeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHis 385
Db 113549 CTCATCTGGCTCTTCCCGCGGGCTGTCTCTCTATTTTGGACCTGAAGCAGGAC 113608

Qy 386 ValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeu 405
Db 113609 GTCTTCGTGATCGTFACTCGTGTGCTGTCTACTTCGCGGCGGTGATGTGTGCTGTG 113668
Qy 406 MetLeuThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerIlePhe 425
Db 113669 ATGCTGACCTGACCCCATCATCTGTGTCGCGCGGTATGCGCTCTCCAAGCTCTTC 113728
Qy 426 AspIleTyrLeuAspPheLys-----ThrSerAspArgIleTyrAlaIle 440
Db 113729 GACGAGTACTCGACTTCTCGACCTGCGCGCGCGCGCGCGCGCGCGCGCG 113788
Qy 441 LysProAlaAla-----LeuLeuAlaLysLeuIleValSerGlySer 454
Db 113789 CGCCCTTCGCGCGCGCGGCGGTTCCTGNAACCTGCTCGGAAGCTCTCTGATCAGGACC 113848
Qy 455 PheIlePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyr 474
Db 113849 TTCCTCTTCTACTCTCTCTCTTGTCTACCATGCACTGCGGTGCTCACTCGACCGCTAC 113908
Qy 475 SerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAsp 494
Db 113909 TCTCGCCCTTCGCTGCTGCTGCTCGCACTCCGACGGCTCCCCCGGCTCATGTAC 113968
Qy 495 AspPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAla 514
Db 113969 GACTACCGGAGCGCTACTACTGCTCGCATGACACCGACCGCGCGCGCGCGCGCG 114028
Qy 515 AlaTrpTrpAspTyrGlyTyrGlnIleGlyMetAlaAspArgThrThrLeuValAsp 534
Db 114029 GCTGTGTGGACTAGGGTACAGATCGCGCGCTGCGCGCGCGCGCGCGCGCGCG 114088
Qy 535 AsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGlu 554
Db 114089 AACACACCTGGAACACACACCATCGCCACTCTCGGCAAGCGCATGGCTCGCGCCAG 114148
Qy 555 GluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGly 574
Db 114149 GACAGGGCTACGAGATCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 114208
Qy 575 GlyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSer 594
Db 114209 GCGTACTCGGCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 114268
Qy 595 GluGlyIleTrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyTyr 614
Db 114269 GAGGGCATCTGCGCTGACGAGATCGCGGAGCGCAACTACTTCAGCGCGCGCGCGAGTAC 114328
Qy 615 ArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyr 634
Db 114329 CGATGGATGCGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 114388
Qy 635 LysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMet 654
Db 114389 CACCGCTTCGCGAGCTCTTCGCGGACACCCGAGCGCGCGCGCGCGCGCGCGCG 114448
Qy 655 IleThrProLeuAspValProProLeuAspTyrPheAspGlnValPheThrSerGluAsn 674
Db 114449 ATCTCGCGCGCGCGCGCGCTCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 114508
Qy 675 TrpMetValArgIleTyrGlnLeuLysAspAspAlaGlnGlyArgThrLeuArgAsp 694
Db 114509 TGGATGTCGCGCATCTACAAGCTCAACACCGCGCGCGCGCGCGCGCGCGCGCG 114568
Qy 695 ValGlyGluLeuThrArg-----SerSerThrLysThrArgArgSerIleLysArgPro 712
Db 114569 GCTAGCGCGCTTCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 114628
Qy 713 GluLeuGlyLeu 716
Db 114629 GCGTGGCGCTC 114640

RESULT 11


```
Db 65786 TTGACGCCCATTTGTTGTTTTCGCTGCCATTCATTTGTTAAATGTTTATGTTTAC 65845
Qy Leu-----AspPheLysThrSerAspArg 436
Db 65846 TTAACGTTGTTGATATTTTCAAGTCTGATTCATTGGAAGATGTTAAACAGATTCTAAG 65905
Qy LysTyAlaIleLysProAlaala----- 444
Db 65906 AATCAAAATCAAAAGTTTCATCGCTGCTCCGCTCAAAATCTTCTAAGTTTATTCAG 65965
Qy LeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyLeuTyLeuPheValPhe 464
Db 65966 TTATTGTCAAAGCAGTTGCTCTATTATTCGTTCACTTTTATCTTTTACTTCTGTAT 66025
Qy HisSerThrTrpValThrArgThrAlaTySerSerProSerValValLeuProSerGln 484
Db 66026 CACTGTACATGGGTACATCAAAATCGGTACTGCTCTCATCGGTGTTTGTAGCTTCAAGA 66085
Qy ThrProAspGlyLysLeuAlaLeuIleAspPheArgGluAlaTyTrpLeuArg 504
Db 66086 AACCCAGATGGTTCTCAAAATATTATTGATGATTATAGAGAAGCTTACTACTGTTGAGA 66145
Qy MetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTyGlyTyGlnIleGly 524
Db 66146 ATGAATACCCCAAGATGCTAAGGTTATGGCATGGTGGGATTATGGCTATCAATCGGT 66205
Qy GlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAla 544
Db 66206 GGGATGGCTGACCGTACTACCTTAGTTGACAACAATACATGGAATACACACACATTGCG 66265
Qy IleValGlyLysValAlaMetAlaSerProGluGluLysSerTyGlnIleLeuLysGluHis 564
Db 66266 ACTGTGGTAAAGCATGTGCTCTCCGAAGATGTGATGATGATGATGATGATGATGAT 66325
Qy AspValAspTyTrpValLeuValIlePheGlyGlyLeuIleGlyPheGlyGlyAspAspIle 584
Db 66326 GATGTGTACTATGATTATCTTTGTTGTTTACTTGGCTTACTTGGCTTACTTGGCTTACT 66385
Qy AsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluIleLysGlu 604
Db 66386 ATAAATCTTATGATGTGTCAGGATTTCCGAAGATATCTGCGCGGATGAAATCCATGAG 66445
Qy ArgTyTrpThrAlaGluGlyGluTyArgValAspAlaArgAlaSerGluThrMet 624
Db 66446 CGTGACTACTTTACAGAAAGAGTCAATACAAAGTTGATGGGATGCTTCCAAAGCAATG 66505
Qy ArgAsnSerLeuLeuTyLysMetSerTyLysAspPheProGlnLeuPheAsnGlyGly 644
Db 66506 AAGACTCTTTGATGTACAAATGTCATATTACAGATACACCGGATTTATTCAATGGTGA 66565
Qy GlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspValProProLeuAsp 664
Db 66566 GATGGTATGGATAGAGTTAGAAACCAAGTATTCTGCTGCAAAATCTCCAGAAATTAGAT 66625
Qy TyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyGlnLeuLysLys 684
Db 66626 GTTCGGGAAGAGCTTTTACTTCTCAAAACCTGGATCGTAAGATCTATAAAGTCAAGAC 66685
Qy AspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArgSerSer----- 702
Db 66686 TTAGACAACTTGTAGATATTACATTTCTGCTGCAAAATTCGATAAATCTTCTTCTCAA 66745
Qy 703 ---ThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArg 717
Db 66746 CAAACCAAGAAAGAGATTTATAAAGAACCATCATATTAGATTAAAGA 66793
```

RESULT 13

CR382128_21/c

WPCOMMENT

Sequence split into 31 fragments LOCUS CR382128 Accession CR382128

Fragment Name

Begin End

CR382128_00 1 110000

```
CR382128_01 100001 210000
CR382128_02 200001 310000
CR382128_03 300001 410000
CR382128_04 400001 510000
CR382128_05 500001 610000
CR382128_06 600001 710000
CR382128_07 700001 810000
CR382128_08 800001 910000
CR382128_09 900001 1010000
CR382128_10 1000001 1110000
CR382128_11 1100001 1210000
CR382128_12 1200001 1310000
CR382128_13 1300001 1410000
CR382128_14 1400001 1510000
CR382128_15 1500001 1610000
CR382128_16 1600001 1710000
CR382128_17 1700001 1810000
CR382128_18 1800001 1910000
CR382128_19 1900001 2010000
CR382128_20 2000001 2110000
CR382128_21 2100001 2210000
CR382128_22 2200001 2310000
CR382128_23 2300001 2410000
CR382128_24 2400001 2510000
CR382128_25 2500001 2610000
CR382128_26 2600001 2710000
CR382128_27 2700001 2810000
CR382128_28 2800001 2910000
CR382128_29 2900001 3010000
CR382128_30 3000001 3066374
Continuation (22 of 31) of CR382128 from base 2100001 (CR382128 Yarrowia lipolytica chr
```

Alignment Scores:
Pred. NO.: 8.14e-197
Score: 2495.50
Percent Similarity: 77.19%
Best Local Similarity: 64.37%
Query Match: 66.35%
DB: 8

US-10-028-384-6 (1-718) x CR382128_21 (1-110000)

```
Qy 6 SerCysValLeuSerValPheGlnThrIleLeuLysLeuValIlePheValAlaIlePhe 25
Db 64802 TCGACAGTCAAGTCGATC---AACCTGCTCTCAAGGTGAGTATTCTGTGACGATTGCT 64746
Qy 26 GlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheLysSerIleIleHisGlu 45
Db 64745 GCGCTGCGGTTTCTGCGCTCTGTTTTCGGTGATTGCGTTTTCAGTCCATCCACGAG 64686
Qy 46 PheAspProTrpPheAsnTyArgAlaThrLysTyLeuValAsnAsnSerPheTyLys 65
Db 64685 TTTGACCCCTTGTTCAACTCCGAGCCTCCAGTATCTGTCACCAACAGTTACTACAAG 64626
Qy 66 PheLeuAsnTrpPheAspAspArgThrTrpTyProLeuGlyArgValThrGlyGlyThr 85
Db 64625 TTCTCAACTGTTTCGACGACAAACAGATGATCCCTCTTGGCGAGTCTACTGGAGTACC 64566
Qy 86 LeuTyProGlyLeuMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeu 105
Db 64565 CTCTACCCGCTCTCATGGTCACCTCTGCTGTCATC---CACAAAGCTGCTCGCAATGGTG 64509
Qy 106 GlyLeuProIleAspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyVal 125
Db 64508 GCGCTGCCCATGTGACATTCGAAACATTTGTGTCATGATGACCCCGCTTCTCGGCCCTC 64449
Qy 126 ThrAlaTrpAlaThrTyGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeu 145
Db 64448 ACGGCTACGCCACTACCTGTTTCCACCAAGTACAGATGAAGCAGGACGAGTCCGCTCTGTTG 64389
Qy 146 AlaAlaGlyPheIleAlaIleValProGlyTyLysSerArgSerValAlaGlySerTy 165
Db 64388 GCCGCTGCTTCATCGGTATTCGCCCGGATACATTTCGATCTGCTGCTGATCCCTAC 64329
```

QY 166 AspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMetPheThrIleLysAla 185
 Db 64328 GATACAGGAGCAATGTCATTTCTGCTCATGATCACCTTGTACACCTGGATCAAGGCC 64269
 QY 186 GlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPheThrPheThrMetVal 205
 Db 64268 CTCAGATGGGTCAGCTTCTGGGGTGTCTGTGCTCCCTTTATTTTACATGCTC 64209
 QY 206 SerAlaThrGlyGlyThrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeu 225
 Db 64208 GCCGCCCTGGGAGATACGTTTATACATAATATGATCCCTTCTCATGTGTGTGCTC 64149
 QY 226 IleLeuMetGlyArgThrSerSerLysLeuThrSerAlaThrThrThrThrThrThrThr 245
 Db 64148 ATCCCTCATGGCCGATACATGACCAATCTACGTGGCTTACGACATGGTATGGCTG 64089
 QY 246 GlyThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAsp 265
 Db 64088 GGTACACTCGCCTCCATGCAATTCCTTTGTGGTTCCTACCCATCGGCTCTGAG 64029
 QY 266 HisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheVal 285
 Db 64028 CACATGGCTGCTCTAGTGCTTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63969
 QY 286 LysGlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuLe 305
 Db 63968 AAGAGCCAGGTCACCCCAAGCAATCAAGTACACCGCTGGTCTGCTGCTGCTGCTGCT 63909
 QY 306 LeuValLeuGlyValValGlyLeuSerAlaLeuThrThrMetGlyLeuIleAlaProThr 325
 Db 63908 ATTGCTATTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63849
 QY 326 ThrGlyArgPheThrSerLeuThrAspThrAspThrAlaLysIleHisIleThrIle 345
 Db 63848 ACCGCCAATCTACTCCCTCTGGATACCGGCTACGCCAAGATCCACATCCCATATT 63789
 QY 346 AlaSerValSerGluHisGlnProValSerThrProAlaPhePhePheAspThrHisPhe 365
 Db 63789 GCCTCTGCTCCGAGCACCAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63729
 QY 366 LeuIleThrLeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHis 385
 Db 63728 CTCATCTGGCTTCTCCCTCTGGTGTCTACTCATCTTCCAGGAGCTGGACAGGAC 63669
 QY 386 ValPheValIleAlaThrSerValLeuCysSerThrPheAlaGlyValMetValArgLeu 405
 Db 63668 ATTTTGTCTGTTGCTACTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 63609
 QY 406 MetLeuThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePhe 425
 Db 63608 ATGCTCACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63549
 QY 426 AspIleThrLeuAspPheLysThrSerAspArg-----LysThr 438
 Db 63548 GACGCTACCGTGACTTCTTACCAAGACTAAGGGGACACTGTTCTGCTTCTTAAGTAT 63489
 QY 439 -----Ala 439
 Db 63488 CAGCCTCTGGTACGAGACCCCTGGTGAAGTGTGTGAGATGACGAGATGACGAGACT 63429
 QY 440 IleLysProAlaAla-----LeuLeuAlaLysLeuIleValSerGlySer 454
 Db 63428 GTCCACAGAGCGGCAAGTTTCCACCTTGGTCTGTTGTTGCCAAGGAGCCGTGATGTCTCC 63369
 QY 455 PheIlePheThrLeuThrLeuPheValPheHisSerThrThrValThrThrThrThrThr 474
 Db 63368 TTCCCTTTTACCTTGGTATGTTTGTCTACATGTCACCTGGGTCTCTTACGCTTAC 63309
 QY 475 SerSerProSerValLeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuAsp 494
 Db 63308 TCTTCCCTCCGTTGTTTCTTGCCTCTCGAATGCCGACGGCTCTCAGCATCTCATTTGAT 63249

QY 495 AspPheArgGluAlaThrThrThrLeuArgMetAsnSerAspGluAspSerLysValAla 514
 Db 63248 GACTACCGAGAGCTTACTTGGCTTGAATGAACACCCCGAGGATGTCGAATCATG 63189
 QY 515 AlaThrThrAspThrGlyThrGlnIleGlyMetAlaAspArgThrThrLeuValAsp 534
 Db 63188 TCTGTGTGGATTCAGGCTTACAGATTGGAGAAATGGCCGACCACTCTTGTGGAT 63129
 QY 535 AsnAsnThrThrAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGlu 554
 Db 63128 AACACACCTGTGAACAAATACCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63069
 QY 555 GluLysSerThrGluIleLeuLysGluHisAspValAspThrValLeuValIlePheGly 574
 Db 63068 GAGTGGCTGCTCCCATCTCTGCGCAGCAGCATGCTGGATTACGTTCTTGTCTGCTG 63009
 QY 575 GlyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuThrMetIleArgIleSer 594
 Db 63008 GGCCATCTTGGTCTCTCGGAGACGACATCAACAGTTCCTGTGATGGTGGNATGCC 62949
 QY 595 GluGlyIleThrProGluGluIleLysGluArgThrPheThrThrAlaGluGlyGluThr 614
 Db 62948 GAGGCAATTTGGCTTCAAGAGGAGGCAAACTACTTCACTCTCTGCTGCTGCTGCTGCT 62889
 QY 615 ArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuThrLysMetSerThr 634
 Db 62888 AGGTTTGACCAACAGGCTCGGCGCATGAAGAACTCGCTCATGTACAAAATGTCTGCT 62829
 QY 635 LysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnMet 654
 Db 62828 TACCGGTTCCGCGAG-----GGCACACAAAAGCTGACCGAGTTCGAAACCACTGG 62778
 QY 655 IleThrProLeuAspValProProLeuAspThrPheAspGluValPheThrSerGluAsn 674
 Db 62777 ATCCCT 62721
 QY 675 ThrMetValArgIleThrGlnLeuLysLysAspAlaGlnIleArgThrLeuArgAsp 694
 Db 62720 TGGCTGTGCTGCTCTATAAGTCAAGGACCTCGATAACCTGGGACCGCATCTGAAGAGC 62661
 QY 695 ValGlyGlu-LeuThr-----ArgSerSerThrLysThrArgArgSe 708
 Db 62660 GCCAAGAGCTTTGACACTGGAAGCCCAAGAGGCGCAAGGCTGGTACTCGAGAG 62601
 QY 708 i 708
 Db 62600 C 62600
 RESULT 14
 SPBC1271/c 40907 bp DNA linear PLN 23-JUN-2003
 LOCUS S.pombe chromosome II cosmid cl271.
 DEFINITION AL034353
 ACCESSION AL034353.2 GI:6912224
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Schizosaccharomyces pombe (fission yeast)
 Schizosaccharomyces pombe
 Schizosaccharomycetes
 Schizosaccharomycetes; Schizosaccharomycetaceae;
 Schizosaccharomycetes;
 1 (bases 1 to 40907)
 REFERENCE
 AUTHORS
 Wood, V., Williams, R., Rajandream, M.A., Lyne, M., Lyne, R.,
 Stewart, A., Scours, J., Peat, N., Hayles, J., Baker, S., Basham, D.,
 Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T.,
 Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P.,
 Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D.,
 Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S.,
 Huckle, E.J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M.,
 Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S., Neil, S.,
 Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neill, S.,
 Pearson, D., Quail, M.A., Rabinovitch, E., Rutherford, K., Rutter, S.,
 Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M.,
 Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R.G.,

Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Art, R., Robben, J., Grymonprez, B., Weltjens, I., Vansireels, E., Rieger, M., Schaefer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehmann, W., Wedler, H., Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Weller, H., Wambutt, R., Purrelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S. J., Xiang, Z., Hunt, C., Moore, K., Hurst, S. M., Lucas, M., Rochet, M., Gallardin, C., Tallada, V. A., Garzon, A., Thode, G., Daga, R. R., Cruzado, B., Jimenez, J. L., Sanchez, M., del Rey, P., Benito, J., Dominguez, A., Revuelta, J. L., Moreno, S., Armstrong, J., Forsburg, S. L., Cerutti, L., Lowe, T., McCombe, W. R., Paulsen, I., Potashkin, J., Shpakovski, G. V., Ussery, D., Barrell, B. G. and Nurse, P.

The genome sequence of *Schizosaccharomyces pombe*
 Nature 415 (6874), 871-880 (2002).

21848401
 11859360
 2 (bases 1 to 40907)
 Lyne, M., Rajandream, M. A., Barrell, B. G., Beck, A., Borzym, K., Klages, S., Langer, I. and Reinhardt, R.

Direct Submission
 Submitted (23-NOV-1998) European Schizosaccharomyces genome sequencing project, Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: pombe@sanger.ac.uk and Max Planck Institut fuer Molekulare Genetik, Ihnestr. 73, D-14195 Berlin, Germany

On Feb 7, 2000 this sequence version replaced gi:3925769.

Notes:
 Details of S. pombe sequencing at the Sanger Institute are available on the World Wide Web
 URL: <http://www.genedb.org/genedb/pombe/index.jsp> (URL)
<http://www.sanger.ac.uk/projects/S.pombe/>
 CDS are numbered using the following system eg SPAC5H10.01c. SP (S. pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c (complementary strand). However, clones may have been reorientated since the original submission, therefore the complementary strand notation may be invalid for strand inference. IMPORANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once.

COMMENT

FEATURES
source

1..40907
 /organism="Schizosaccharomyces pombe"
 /mol_type="genomic DNA"
 /strain="972h."
 /db_xref="taxon:4896"
 /chromosome="II"
 /map="1L"
 /clone="cosmid c1271"
 1307..3367
 /gene="SPBC1271.15c"
 1307..3367
 /gene="SPBC1271.15c"
 /note="translation initiation factor if-2 (predicted); localization mitochondrial (predicted); non-essential (PMID 12618370); similar to S. cerevisiae YOL023W"

/codon_start=1
 /protein_id="CAA22205.1"
 /db_xref="GI:3925784"
 /db_xref="GOA:O59683"
 /db_xref="Swiss-Prot:O59683"

/translation="MAFLSACTSRVHVNAGFTITKFPVRNVHKLVSYHISPLASRF SLADISVKNHNGPOLRLQTHPHSSSPHATKGRFPKSTPLTLPAVLSVASFA NLLQIPTRILRLNRLGKIDVFEEITVEYSLIAEYDFDVQAAIYAPSPASA NKKSDGLRPVPTLMGVHDHKTLLDPAFRKSTIASTEGHITQIKGAFVFPDKG SKFTFLDTPGHMAFEAMKRGANIADIVLVVAGDGVKQPTVEAIKHQIADVFV VALTKSPFGPIHKIYEQLNNGIQVEALGGTQIIPISAKTGKGIPELAAITLRA BIMEIRASPFQFQWIVSVSSVTGVGSSATVVVXRGITGVKGMVLVAGSWCKVRSLV LDKSKIKOVLPGQAAQVYGKDLPIAGDLAYEVSESEAKRILSDIYRQSEONQFYE LAESCNEORSAKSGGAAIQEETSVSNSFNIAKCEDTGSMEALSDYLPKQFV KVKSRVLTGVTGPTVDIERATSDAILISFGVSVPKATPFLAEKHNTKLLFHNVIY HLMDVRLKFLALRPLPILVORVGEAILISALFDIKARVHVAGCRVNTGNTIEKSH IRLVNDKILWSGEIDSKLKLBKEVTSIKKRGCGILLNFDIETGDKLQTPVBEYK

gene

CDS

misc_feature

misc_feature

misc_feature

gene

CDS

misc_feature

misc_RNA

gene

CDS

PPDF"
 1811..2815
 /gene="SPBC1271.15c"
 /note="pfam match to entry PF00009 GTP_EFTU, Elongation fac tor Tu family"
 1814..2815
 /gene="SPBC1271.15c"
 /note="Match to GTP_EFTU, Score 114.20"
 3215..3283
 /gene="SPBC1271.15c"
 /note="PS01176 Initiation factor 2 signature"
 complement(3720..5057)
 /gene="SPBC1271.14"
 complement(3720..5057)
 /gene="SPBC1271.14"
 /EC_number="2.3.1.35"
 /EC_number="2.3.1.1"
 /note="amino-acid N-acetyltransferase (predicted); glutamate N-acetyltransferase (predicted); involved in ornithine biosynthesis (predicted); involved in arginine biosynthesis (predicted); amino acid N-acetyltransferase activity; glutamate N-acetyltransferase activity; localization mitochondrial matrix (predicted); non-essential (PMID 12618370); similar to S. cerevisiae YMR062C
 PATH:map00220
 aspect=P; term=ornithine biosynthesis; GOid=GO:0006592; evidence=ISS; db_xref=GO:unpublished; with=SPTR:Q04728; date=20030616
 aspect=P; term=arginine biosynthesis; GOid=GO:0006526; evidence=ISS; db_xref=GO:unpublished; with=SPTR:Q04728; date=20030616
 aspect=F; term=amino acid N-acetyltransferase; GOid=GO:0004042; evidence=ISS; db_xref=GO:unpublished; with=SPTR:Q04728; date=20030616
 aspect=F; term=glutamate N-acetyltransferase; GOid=GO:0004358; evidence=ISS; db_xref=GO:unpublished; with=SPTR:Q04728; date=20030616
 aspect=C; term=mitochondrial matrix; GOid=GO:0005759; evidence=ISS; db_xref=GO:unpublished; with=SPTR:Q04728; date=20030616
 /codon_start=1
 /protein_id="CAA22204.1"
 /db_xref="GI:3925783"
 /db_xref="GOA:O94346"
 /db_xref="TrEMBL:O94346"
 /translation="MQTKIFARLAGIGRARFVGTMPDKTKFSKSTYPOQSILNGIA SGVANGKDLALFSEFCNAARFTKNAFQAPQVSRQINGCGKDIHCYVFN SGNVATVGGGLMDAQLITAEADNLTRPHWTENSEEFPSSLVNMTVIGVRLKD KIOSGLEHVEDLGSTHEYWMRAAEICTTDFPKLVSRSLIAGKVRIAGFAKAG MINPNLATLLGLEVDAPISVDVRSILRHAINNSFNISIDGDTSTNDTIAFLANGA AGSEITKSSPAYKEIRDAVDTIAQLAKLVVDGEGATKFTVTVVRGARSKDAALV ASTISNALKVTAFFGEDANWGRILCAVGSGAAVNPATTVSFIPADTEPLKLVN GEPQNVDEFRASILLSQDELTVDDVDLCGCPYAKTNWTCDFSVDYVRINADYS"

complement(3723..4946)
 /gene="SPBC1271.14"
 /note="Match to Argu, Score 840.54"
 complement(6265..6931)
 /note="SPMEIRNA D31852 Fission yeast DNA for meiRNA, complete sequence"
 complement(join(7889..8438,8502..8575))
 /gene="SPBC1271.13"
 complement(join(7889..8438,8502..8575))
 /gene="SPBC1271.13"
 /note="mitochondrial ribosomal protein L17; essential (PMID 12618370); similar to S. cerevisiae YJL063C"
 /protein_id="CAA22203.1"
 /db_xref="GI:3925782"
 /db_xref="GOA:O94345"
 /db_xref="TrEMBL:O94345"
 /translation="MTKLYFRKLGRTSARHQRALLNLVTSLVKHSQTQTTWAKAEA ORFAEKLLITNAKRNPNRNRKLAEGWVFEKSTTLKKVDFVLVPRYNGRCGYTRLLK

LPRSTNDPMGVLEFVDGPKDIRHMTAKTVGICLQAKALPTLRKHVLLFRK
 NGXARFDQVQKESSEHARLKEDHEDEKTVKDWKRDPIPRFYI
 /genes="SPBC1271.13"
 /notes="Match to Ribosomal L17, Score 114.95"
 /notes="SPBC1271.13"
 /genes="SPBC1271.13"
 /notes="cctaataaagggaacag, splice branch and acceptor"
 /complement (8439, .8454)
 /genes="SPBC1271.13"
 /complement (8496, 8501)
 /genes="SPBC1271.13"
 /notes="gtagct, splice donor sequence"
 /complement (8899, .9327)
 /notes="mRNA from AH016222"
 /complement (9328, .9869, 9929, .10241, 10424, .10546,
 10690, .10768, 10941, .10966, 11042, .11054, 11105, .11175)
 /genes="keel"
 /notes="synonym: SPBC1271.12"
 /complement (9328, .9869, 9929, .10241, 10424, .10546,
 10890, .10768, 10941, .10966, 11042, .11054, 11105, .11175)
 /genes="keel"
 /notes="oxysterol binding protein (predicted); involved in
 ergosterol biosynthesis (predicted); similar to S.
 cerevisiae YHR001W and YHR003W and YOR237W and YPI145C;
 similar to S. pombe SPBC646.08c and SPBC354.07c and
 SPCC2396.01; localization golgi (predicted); non-essential
 (PMID 12618370)"
 /codon_start=1
 /protein_id="CAA22202.1"
 /db_xref="GI:3925781"
 /db_xref="GOA:O74178"
 /db_xref="TrEMBL:O74178"

Alignment Scores:

Pred. No.: 9,64e-176 Length: 40907
 Score: 2237.00 Matches: 425
 Percent Similarity: 72.84% Conservative: 114
 Best Local Similarity: 57.43% Mismatches: 167
 Query Match: 59.48% Indels: 34
 DB: 8 Gaps: 5

US-10-028-384-6 (1-718) x SPBC1271 (1-40907)

QY 7 CysValLeuSerValPheGlnThrIleLeuLysLeuValIlePheValAlaIlePheGly 26
 DB 37681 TGCATAATAGCTGTATTAGCATTGGTAAATCAAGTATCTTTAGT----- 37634
 QY 27 AlaAlaIleSerArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPhe 46
 DB 37633 GCITCTGCTCTTCGAGGCTTTTCTGTCATTCTGTTACGAGTCCCATTTATTCATGAATT 37574
 QY 47 AspProTrpPheAsnTyrAlaThrLysTyrLeuValAsnAsnSerPheTyrIlyPhe 66
 DB 37573 GATCCITGGTCAATTCGAGCTTCCAAATATTGGTGAACAAAGGTTTATTAACCTT 37514
 QY 67 LeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArgValThrGlyClyThrLeu 86
 DB 37513 TTAATTTGTTTCATGAAGAAGTTGTTACCCGTTGGTCGTTAGCGGGTGTACTTTG 37454
 QY 87 TyrProGlyLeuMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGly 106
 DB 37453 TACCAGAGACTTATGGTCAGCTGTGTATTATTTTCAAAAGTTTAA--CATCTTTTAAAG 37397
 QY 107 LeuProIleAspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThr 126
 DB 37396 ATTAACGTGAACATTCTGTATGTATGTTTACTTGCCTGCTTTCTCTGGAATCACT 37337
 QY 127 AlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAla 146
 DB 37336 CGGATTCTACCTATTATCTGGCTAGAGAAATGAAGAAGTGATGCGCTTTTAGCT 37277
 QY 147 AlaGlyPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAsp 166
 DB 37276 GCCGCAITTTATGGGTATTGCTCTGGATACACCTCCGTCGCTCGCTGGTTCCTACGAT 37217

QY 167 AsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMetPheThrIleLysAlaGln 186
 DB 37216 AATGAAGCAATTCCTATTACCTTTTGATGCAACAGTTGCTTTGGGATCAAGGAGTG 37157
 QY 187 LysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSer 206
 DB 37156 AAGTCTGGCTCTCTTTCTGGGGTCCCTGCACAGATTGCTACTTCTATATGTTAACT 37097
 QY 207 AlaTrpGlyGlyTyrValPheIleThrAsnLeuLeuProLeuHisValPheLeuLeu 226
 DB 37096 GCGTGGGGTGTATGATTCATCAACAATGATACCTTTACAGTATTGTTCTCTTA 37037
 QY 227 LeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGly 246
 DB 37036 CTTATGGGTGCTATCTACTAGCAATTTATACATCTCTTACACAACATCTATGTTATGGA 36977
 QY 247 ThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHis 266
 DB 36976 AGCTGGCTTCATCAAGTTCGGTTTGGTTTCCACCGGTGCGACTGATGAGCAT 36917
 QY 267 MetAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLys 286
 DB 36916 ATGTCGGTTTAGGAGTGTGGCTGTACAGCTTTTTCATCTACAATTTATGTTAAA 36857
 QY 287 GlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuLeuLeu 306
 DB 36856 GGTCTAGTTTCATCCAGCAATTCAAATCTATTGTTTGGTTTGGTTTGGTTTCTTAGTG 36797
 QY 307 ValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThr 326
 DB 36796 GGTCTAGCAACAGTCTCTTTTGTCTTATCTCAACAGGTGTATCGCTCTCTTGACA 36737
 QY 327 GlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleAla 346
 DB 36736 GACGCTTCTATCTCTTTGGATACAACTPACCCAGATTCATATCTCTATCTGCT 36677
 QY 347 SerValSerGluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeu 366
 DB 36676 TCGGTATCAGAACATCATCGCTCTCTTGTGGAGTTCGTTGTTCTTTGATCTTCAATTTTG 36617
 QY 367 IleTrpLeuPheProAlaGlyValPheLeuLeuLeuAspLeuLysAspGluHisVal 386
 DB 36616 ATTGGTTATTGCCAGTTGGTTTACTTGTGTTTCAAGAACTTCGTAATGAACATGTC 36557
 QY 387 PheValIleAlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMet 406
 DB 36556 TTTATTATATATATCTCTTAGGAACATATTTTGTGGTGTGATGGTTCGTTTGGTT 36497
 QY 407 LeuThrLeuTrpProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAsp 426
 DB 36496 TTAACCTTAACCTCTGTGTTTGTATGAGTCTGCTGTAGCAATTCCTCTTTTAGAC 36437
 QY 427 IleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIleLysProAla----- 443
 DB 36436 ACATATATGGTCTCTGAAGTTGAAGAGGACAAAGTGACGGAAGAGCGCTTCAGCCAAA 36377
 QY 443 ----- 443
 DB 36376 TCTAAGAACAGAAAGGATTCTTCTATCTCTTATGTTTCTTCTGCTCTCAAAAAAT 36317
 QY 444 -----AlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeu 459
 DB 36316 ATTGGAATTTACAGTTTCTTCCAGAGTATTAGTCAATTTCTCTACCGCATATTCTCTA 36257
 QY 460 TyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerVal 479
 DB 36256 ATAAGTTTGTATTATCTATTCAGTTGGTGTACTCTAATGTTACTTCTCCCTACCGTG 36197
 QY 480 ValLeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuAspPheArgGluAla 499
 DB 36196 GTTTTGTCTACCGTGTAAACGAGTGTAGTTTAAATGATATATTGATGACTTCCGTAAGCT 36137

```

QY 500 TyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTyr 519
Db 36136 TATGACTGCTTCGTAGAACACTCTCTATGACAAAGGTATGAGTGTGGGATAT 36077
QY 520 GlyTyrGlnIleGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsn 539
Db 36076 GGTACCAAAATTCGTGATGCTGATGCTATCTAGTCGACAAACATACGTGGAAC 36017
QY 540 AsnThrHisAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGlu 559
Db 36016 AACACACATTTGCCACAGTTCGAAAGCAATGCTTCACCTGAAGAAAGCTTACCT 35957
QY 560 IleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPhe 579
Db 35956 ATCTCTCGTAACACAGATCTGATATATCTTATATATATATATGCTGCTACTCTGGATAC 35897
QY 580 GlyGlyAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpPro 599
Db 35896 AGCAGCGACGACATGACAAAGTCTCTTGATGATGCGAATCTCAGGATATGCGCC 35837
QY 600 GluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArg 619
Db 35836 GATGAATAGTAGACGTAACCTTTTACTCTAATGAGAATATCGAACTGACGATGCG 35777
QY 620 AlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGln 639
Db 35776 GCTACTCCACATTCGCTGAGCTTATATATAGATGTCATATACCGTGTGGA 35717
QY 640 LeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAsp 659
Db 35716 CTTTTCCCTCCCAATCAAGGATATGACCGTGTGCGAATCAAAACCTACCATCGAAGAT 35657
QY 660 ValProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIle 679
Db 35656 ---CTCACTATTACTATCGAAGACATCTACTCCGTTCACTATTAGTTCGTTG 35600
QY 680 TyrGlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThr 699
Db 35599 TATAAGGTTAAGAAACCGGATACACTTGGACGCGATTGAAACAAAGTGACATATTGAA 35540
QY 700 ArgSerSer---ThrLysThrArgSerIleLysArgProGluLeuGlyLeuArgVal 718
Db 35539 GAAGCAAGAAAGAGCTCGCGCTGCTCGAAACTCAAGAAATCTCTTGGAGGTT 35480

RESULT 15
AB015232
LOCUS
DEFINITION
Schizosaccharomyces pombe stt3+ gene for oligosaccharyltransferase
subunit, complete cds.
ACCESSION
AB015232
VERSION
AB015232.1 GI:4586802
KEYWORDS
oligosaccharyltransferase subunit; stt3+.
SOURCE
Schizosaccharomyces pombe (fission yeast)
ORGANISM
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
REFERENCE
1 (sites)
Yoshida,S., Matsuura,A., Merregaert,J. and Anraku,Y.
Schizosaccharomycetes pombe stt3+ is a functional homologue of
Saccharomycetes cerevisiae SRT3 which regulates
oligosaccharyltransferase activity
Yeast 15 (6), 497-505 (1999)
JOURNAL
MEDLINE
PUBMED
10234787
REFERENCE
2 (bases 1 to 5048)
Yoshida,S.
Direct Submission
Submitted (04-JUN-1998) Satoshi Yoshida, Kirin Brewery Co., Ltd.,
Central Laboratories for Key Technology; 1-13-5, Fukuura
Kanazawa-ku, Yokohama, Kanagawa 236-0004, Japan
(E-mail:satoshi@kitrin.co.jp, Tel:81-45-788-7360,
Fax:81-45-788-4042)

```

FEATURES

```

Location/Qualifiers
1..5048
/organism="Schizosaccharomyces pombe"
/mol_type="genomic DNA"
/db_xref="taxon:4896"
join(1203..1302,1362..3511)
/gene="stt3+"
join(1203..1302,1362..3511)
/gene="stt3+"
/note="similar to S. cerevisiae, mouse and human
homologue"
/codon_start=1
/product="oligosaccharyltransferase subunit"
/protein_id="BAA76479.1"
/db_xref="GI:4586803"
/translators="MANSARI"

```

```

IRYESIHEDEDPWFNFRASKLIVLVEQSYNPLWFDERSWYPIGRVAGGTLYGLMVTIS
GIIFKHLJLRKNNINDVCVLLAPAFSGITATATYLLARELKSACGLAFAFMGIA
PGYTSRVAGSYDNEAITLMTSLMTKAVKSSGFWGACTGLLYFVWANGGY
VFITNMPLHVFVLLMGRYTSKLYATYTYVIGTASMQVFPVGFQVSTSEHMSA
LGVFGLQLFAPYNYKGLVSRQFQILIRFALVCLVGLATVVLFSALSTGVIAWPTG
RFYSLMDVNTAKIHIPILASVSEHQPTWSSLPFDLOFLWLLPVGVLCFKELENEH
VFIIYPLVGTFCGVVVRVLVLTPTCVCAAAVATSTLLDTWGVPEVEDKVSBEAA
SAKSNKKGSSILSPFTSGKNIGLYSLLSVLVLSSTAYELIMVHVHSSVSNAY
SSPTVVLSTVLDGSLMIDDFEADWLRNTPYDTKVMWHDVGYQIAGMNRITL
VDNMNTHIATVGMKSPSEKAYPILRKHDDVIXILLYGGTGLGYSDDMKELWM
IRISQGLWPEIDVERNFFTPNGEYRDDAATPTMRESLLYKMSYGAWLFPNGOYD
RARNOKLPEDKPOLFTIEEAFITVHVLVLYKKPDTLGRDLKQVTLFEGKRKKA
VLQKLTAKFL
1303..1361
/gene="stt3+"

```

```

intron
ORIGIN

```

```

Alignment Scores:
Pred. No.: 6.88e-176 Length: 5048
Score: 2224.50 Matches: 424
Percent Similarity: 72.97% Conservative: 116
Best Local Similarity: 57.30% Mismatches: 166
Query Match: 59.15% Indels: 35
DB: 8 Gaps: 5

```

```

US-10-028-384-6 (1-718) x AB015232 (1-5048)

```

```

QY 7 CysValLeuSerValPheGlnThrIleLeuLysLeuValIlePheValAlaIlePheGly 26
Db 1316 TGCATAATAGCTGATTAGCAGTGTGTTAATCAAGTATCTTTTAGTG----- 1363
QY 27 AlaAlaIleSerSerArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPhe 46
Db 1364 GCTTCTGCTCTTCGAGGCTTTTCTGTCATTCGTTAGAGTCCATTAATCATGAATTT 1423
QY 47 AspProTrpPheAsnTyrArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPhe 66
Db 1424 GATCCTTGGTTCATTTCCGAGCTTCCAAAATATTGCTGGAACAGGTTTTTATAACTTT 1483
QY 67 LeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeu 86
Db 1484 TTAATTTGGTTTGAAGAAGATTTGATCCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1543
QY 87 TyrProGlyLeuMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGly 106
Db 1544 TACCAGGACTTATGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1600
QY 107 LeuProIleAspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThr 126
Db 1601 ATTAACGTGAACATCCGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1660
QY 127 AlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAla 146
Db 1661 GCGATTCTACCTATTATCTGCTAGAGAAATGAAAGATGATGATGATGATGATGATGATG 1720
QY 147 AlaGlyPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAsp 166

```

Db 1721 GCCGCATTATGGGTATTTGCTCTGGATACACCTCCGTTCCGTCCTGCTTCTTACGAT 1780
QY 167 AenGluAlaIleAlaIleThrLeuLeuMetValThrPheMetPhePheIleLysAlaGln 186
Db 1781 AATGAAGCAATTCGTTATACCTTTTCATGTCACGTTTCTTGGGATCAAGCAGTG 1840
QY 187 LysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhePhePhePhePhePhePhe 206
Db 1841 AAGTCTGGCTCTCTCTCTGGGGGCTGCTGACAGGATGCTCTACTTCTATATGGTAAT 1900
QY 207 AlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeulle 226
Db 1901 GGTGGGGTGGTATGATGATTCATCACAACATGATACCTTTACACGATTTTGTCTCTA 1960
QY 227 LeuMetGlyArgTyrSerSerLeuLeuTyrSerAlaTyrThrThrTrpTyrAlaIleGly 246
Db 1961 CTTATGGGTGCTATCTAGCAAAATATACATTCCTTACACACATCTATGTTATGGA 2020
QY 247 ThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHis 266
Db 2021 AGCTGGCTTCTATGCAAGTCCGTTTGTGGTTTCCACCCGTCGACCTAGTACGAT 2080
QY 267 MetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLys 286
Db 2081 ATGCCGCTTTAGAGTGTGGCTGTTACAGCTTTTGGCATTCACATTAATGTTAAA 2140
QY 287 GlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeu 306
Db 2141 GGTCTAGTATTCATCAAGCAATTCCAAAATATCTATTGCTTTGCTTGGCTTAGTG 2200
QY 307 ValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProThr 326
Db 2201 GGTCTAGCAACAGTCGCTTTTGTCTTATCTTACACAGGTGTATCGCTCTTGGACA 2260
QY 327 GlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleLeala 346
Db 2261 GGACGTTTCTATCTCTTGGGATACAAACTAGCCCAAGATTCATATCTCTATCTGCT 2320
QY 347 SerValSerGluHisGlnProValSerTrpAlaPhePhePhePhePhePhePheLeu 366
Db 2321 TCGGTATCAGAACATCAGCTCTCTACTTGGAGTTCGTTCTTCTTGTATCTCAATTTTG 2380
QY 367 IleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysIleHisVal 386
Db 2381 ATTTCGTTATTCGCAAGTGTGGTGTATCTTGTGTTTCAAGAACTTCGTAATGAACATGTC 2440
QY 387 PheValIleAlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMet 406
Db 2441 TTATATTATATATCTCTTCTTAGGAACATATTTTGTGTGTGATGTTCTGTTTGGTT 2500
QY 407 LeuThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAsp 426
Db 2501 TTACCTTAACTCTCTGTTTGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2560
QY 427 IleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIleLysProAla 443
Db 2561 ACATATATGGTCTCTGAAGTTGAAGAGGACAAAGTGAAGGAGGAGGAGGAGGAGGAGG 2620
QY 443 2621 TCTAAGAACAAAGGATTTCTCTATCTTCTTAGTTTCTTCTCTCTCTCTCTCTCTCTCTCT 2680
QY 444 2681 ATTGGAATTTACAGTTTGTCTTCCAGAGTATAGTCATTCCTCTACCGCATATTTCTCTA 2740
QY 460 TyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerVal 479
Db 2741 ATAATGTTTGTATTATCATTCAGTGGGTGACTTCTTAATGCTTACTTCTCCCTACCGTG 2800
QY 480 ValLeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuLeuLeuLeuLeuLeuLeu 499
Db 2801 GTTTTGTCTACCGTGTAAACGATGGTAGTTTAAATGATATATATGATGACTTCCGTGAAGCT 2860

QY 500 TyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaIleTrpTrpAspTyr 519
Db 2861 TATGACTGGCTTCGTAGAACACTCTTATGACAAAGGTTATGAGTTGGTGGGATAT 2920
QY 520 GlyTyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsn 539
Db 2921 GGTACCAAAATGCTGGTATGCTGATGCTATTTACTTTAGTCGACAACTACGTCGAAC 2980
QY 540 AsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGlu 559
Db 2981 AACACACATATTCACAGTTGGAAAAGCCATGCTTTCACCTGCAAGAAAAGTACCT 3040
QY 560 IleLeuLysGluHisAspValAspTyrValIleuValIlePheGlyGlyLeuIleGlyPhe 579
Db 3041 ATCTCTCCGTAAACAGATGTTGATATATCTTATATATATATGTTGTTACTCTTGTATAC 3100
QY 580 GlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpPro 599
Db 3101 AGCAGCGACGACATGAACAGTTCCTTTGGATGATCCGAAATTTCTCAGGGAATATGGCCC 3160
QY 600 GluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGlyTyrArgValAspAlaArg 619
Db 3161 GATGAATAGTAGACGCTAACTTTTACTCTTAATGAGAATATCGACTGACGATGCG 3220
QY 620 AlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGln 639
Db 3221 GCTACTCCCACTATGCTGAGTCTTTATATATAAGATGTATATCATCAGGTGCTTGGAAA 3280
QY 640 LeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAsp 659
Db 3281 CTTTCCCTCCCAATCAGGATATGACCGTCTCGCAATCAAAACTACCATCGAAGAT 3340
QY 660 ValProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIle 679
Db 3341 ---CCTCAACTATTTACTATCGAAGAAGCATTCCTACCGTTTCATCATTTAGTTCGTTTG 3397
QY 680 TyrGlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeu--- 698
Db 3398 TATAAGGTTAAGAAACCGATACACTTGGACGGGATTTGAACAAGTGACATT-ATTGCA 3456
QY 699 ThrArgSerSerThrLysThrArgSerIleLysArgProGluLeuGlyLeuArgVal 718
Db 3457 AGAAGGCAAAAGAAAGATCCGCGCTCTGCAAAAATAACGAAATTCCTTTGAGGGT 3516

Search completed: December 15, 2004, 05:47:27

Job time : 8915.54 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:28:05 ; Search time 827.493 Seconds
(without alignments)
4554.822 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 3761

Sequence: 1 MGSDRSCVLSVFQILKLV.....TRSTKTRRSIKRPELGLRV 718

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10028384/runat.14122004.131518.8007/app_query.fasta.1.3740
-DB=N_Geneseq_23Sep04 -QPMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USRF=US10028384 @CGN 1 1 1986 @runat.14122004.131518.8007 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQURY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_23Sep04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2002bs:*
8: Geneseq2003as:*
9: Geneseq2003bs:*
10: Geneseq2003cs:*
11: Geneseq2003ds:*
12: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3761	100.0	2733	10	Add94787 Yeast STT
2	3488	92.7	2000	10	ACC61293 Gene sequ
3	3488	92.7	2000	10	Adk63659 Disease t
4	2738	72.8	2256	6	Abt20789 Aspergill
5	2428.5	64.6	2232	8	Abt20789 Aspergill
6	2301.5	61.2	2466	10	AdB69900 C. neofo

7	2364.5	60.2	2603	8	ABT20191
8	2364.5	60.2	3969	8	ABT17781
9	2364.5	60.2	4603	8	ABT19595
10	2090	55.6	1848	8	ABT18369
11	2042	54.3	1369	8	ABT18375
12	2032.5	54.0	2882	10	ADB69539
13	2032.5	54.0	4738	10	ADB69178
14	2027.5	53.9	2417	10	ADD94789
15	2027.5	53.9	2699	4	ABL13247
16	2013.5	53.5	2710	10	ADD94785
17	2012.5	53.5	2481	10	ADD94783
18	1955.5	52.0	4922	4	ABL13246
19	1945	51.7	3094	10	ADD94791
20	1920	51.1	2472	10	ADD94793
21	1920	51.1	2472	10	ADD94793
22	1920	51.1	2760	5	ABV24502
23	1902.5	50.6	2855	4	ABL02795
24	1655	44.3	6153	4	ABL02794
25	1259	33.5	1371	12	ADP28508
26	1186.5	31.5	1664	4	AAK94164
27	1186.5	31.5	1664	12	ADL30661
28	1156	30.7	1543	2	AAK85055
29	1156	30.7	1543	8	ACD18981
30	1156	30.7	1543	12	ADG78372
31	1156	30.7	1543	12	ADN60663
32	1154	30.7	1209	4	AAH33264
33	1154	30.7	1209	6	ABL89850
34	739	19.6	1114	4	AAH99794
35	719.5	19.1	2547	4	AD08315
36	719.5	19.1	2660	6	ABO54750
37	718.5	19.1	2510	5	AAK93772
38	718.5	19.1	2537	4	AD08289
39	718.5	19.1	2546	2	AAV44866
40	718.5	19.1	2546	5	AAK98463
41	711.5	18.9	2284	4	AAH18021
42	675	17.9	381	6	ABN23953
43	664	17.7	787	4	AAH07526
44	637.5	17.0	2953	4	ABL18224
45	620	16.5	764	5	AAK93368

ALIGNMENTS

RESULT 1

ADD94787

ID ADD94787 standard; DNA; 2733 BP.

XX

AC ADD94787;

XX

DT 29-JAN-2004 (first entry)

XX

DE Yeast STT3 gene sequence.

XX

KW source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cystostatic; immunosuppressive; antineoplastic; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide; Yeast;
KW gene; ds; STT3.

XX Saccharomyces cerevisiae.

XX WO2003054008-A2.

XX 03-JUL-2003.

XX 18-DEC-2002; 2002MO-CA001967.

XX 20-DEC-2001; 2001US-00028384.

XX (COMP-) COMPATIGENE INC.

PA

XX Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 DR P-PSDB; ADD94788.
 XX
 PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.

XX Disclosure; SEQ ID NO 5; 66pp; English.

XX This invention relates to a novel isolated or purified human protein, termed
 CC source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antitumor therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestinal cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the yeast SPM3 gene which is related to the
 CC invention. Note: this sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.

XX SQ Sequence 2733 BP; 712 A; 597 C; 583 G; 841 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2733
 Score: 3761.00 Matches: 718
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADD94787 (1-2733)

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 DB 219 ATGGGATCCGACCGGTGGTGTGTTGTCGTGTTTCAGACCATCTCAAGCTCGTCATC 278
 QY 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 DB 279 TTGGTGGCGATTTTGGGGTGGCATATCATCACGTTTGTTCAGTCATCAAAATTGAG 338
 QY 41 SerIleIleHisGluPheAspProTrpPheAsnThrArgAlaThrIysTrpLeuValAsn 60
 DB 339 TCTATTATCCATGAATTCGACCGCTGGTTCATATATAGGGCTACCAAAATATCTCGTCAAC 398
 QY 61 AsnSerPheTrpLysPheLeuAsnTrpPheAspArgThrTrpTrpProLeuGlyArg 80
 DB 399 AATTCGTTTACAAAGTTTGAAGTGGTGTGACGACCTAGTGGTACCCCTCGAAGG 458
 QY 81 ValThrGlyGlyThrLeuTrpProGlyLeuMetThrThrSerAlaPheIleTrpHisAla 100
 DB 459 GTTACTGGAGGACATTTATATCTGTTTGTATGACGACCTAGTGGTTCATCTGGCAGCC 518
 QY 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 DB 519 CTGGCAACTGGTTGGGCTTGGCCATTTGACATCAGAAACGTTTGTGTCTATTGGCCCA 578
 QY 121 LeuPheSerGlyValThrAlaTrpAlaThrTrpGluPheThrLysGluIleLysAspAla 140
 DB 579 CTAATTTCTGGGTCACCGCTGGGCTTACGAATTTACGAAGAGATTAAGATGCC 638
 QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTrpIleSerArgSer 160

DB 639 AGCGCTGGGCTTTTGGCTGCTGTTTATPAGCAATGTCCCGGTTATATATCTAGATCA 698
 QY 161 ValAlaGlySerTyAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 DB 699 GTGCGGGGTCTTACGATTAATGAGCCATTGCGCATTAATGCTACATTAATGCTACATG 758
 QY 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 DB 759 TTTTGATTAAGGCCCCAAAAGACTGGCTCTATATGACGACGCAACGCTGTGCGACTTATTC 818
 QY 201 TyrPheTyMetValSerAlaTrpGlyGlyTrpValPheIleThrAsnLeuIleProLeu 220
 DB 819 TACTTCTACATGGTGTGGCTTGGGTGGATACGTTTCATCCCAACTGATGATCCCATC 878
 QY 221 HisValPheLeuLeuIleLeuMetGlyArgTrpSerSerLysLeuTySerAlaTrpThr 240
 DB 879 CATGCTCTTTTGTCTGATTTGATGGGCAGATATCGTCCAAACTGATTTCTGCTACACC 938
 QY 241 ThrTrpTyAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 DB 939 ACTTGTGACGCTATTGGAACCTGTGCATCCATGCGATCCCATTTGTCGGTTTCTACCT 998
 QY 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 DB 999 ATCAGGCTTAACGACACACATGGCCGATTTGGGTGTTTTCGGTTGATTGATTCAGATTC 1058
 QY 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetVal 300
 DB 1059 TTCGGTGACTTCGTGAAGGGCCAAATCAGCAGACTAAGTTTAAAGTCATCAATGATGGT 1118
 QY 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTrpMetGly 320
 DB 1119 TCTCTGTTTGTGATCTTGTGCTTGTGGTGGGACTTTCGCTTGACCTATATGCGG 1178
 QY 321 LeuIleAlaProTrpTrpThrGlyArgPheTySerLeuTrpAspThrAsnTrpLysIle 340
 DB 1179 TTGATTCGCCCTTGGAGCTGGTAGATTTTATTCGTTATGGGATACCAACATACCAAGATC 1238
 QY 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProAlaPhePhe 360
 DB 1239 CACATTCCTATCATTCGCTCCGTTTCCGAACATCAACCCGTTTCGTTGGCCGCTTCCTC 1298
 QY 361 PheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 DB 1299 TTTGATACCCACATTTTGTATCTGGCTATTCGCCGCGGTGATATTCCTACTATTCTCTGAC 1358
 QY 381 LeuLysAspGluHisValPheValIleAlaTySerValLeuCysSerTyTrpPheAlaGly 400
 DB 1359 TTGAAGACGACGACGCTTTTGTATCGCTTACTCCGTTCTGTGTTCTGTTTGGCGGT 1418
 QY 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
 DB 1419 GTTATGTTAGTATGATGTTGACTTTGACACACCATCATCTGTGTGCGCGCGCTCGCA 1478
 QY 421 LeuSerLysIlePheAspIleTyLeuAspPheLysThrSerAspArgLysTyAlaIle 440
 DB 1479 TTGTCCAAAGATATTGACATCTACCTCGATTTTCAAGACAAAGTACCCGCAAAATACGCCATC 1538
 QY 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyLeuTy 460
 DB 1539 AAACCTGGGACCTACTGGCCAAATGATTGTTTCCGGATCATCTCTTTTATTGTAT 1598
 QY 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTySerSerProSerValVal 480
 DB 1599 CTTTTCGCTCTCCATTTACTTGGGTAAACAAAGAACTGCATACTCTCTCTCTCTCTCTCT 1658
 QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTy 500
 DB 1659 TTGCATCACAAACCCAGATGTTAAATTTGGGTTGATCCGACCTTCAGGAAGGATAC 1718
 QY 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTyGly 520

1921 TTCAATGGTCCCAAGCCATGACAGAGTCCGTCACAAATGATACACCATAGACGTC 1980
 661 ProProteinasePyrPhe 666
 1981 CCACCATAGACTACTTC 1998

RESULT 4
 ABZ32036
 ID ABZ32036 standard; DNA; 2256 BP.
 XX
 AC ABZ32036;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE Candida albicans essential gene SEQ ID NO 6323.
 XX
 KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KW signal transduction; DNA replication; cell division; growth;
 KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
 XX
 OS Candida albicans.
 XX
 PN WO200253728-A2.
 XX
 PD 11-JUL-2002.
 XX
 PF 26-DEC-2001; 2001WO-US049486.
 XX
 PR 29-DEC-2000; 2000US-0259128P.
 PR 20-FEB-2001; 2001US-00792024.
 PR 22-AUG-2001; 2001US-0314050P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 PI
 PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
 XX
 DR WPI; 2002-566694/60.
 DR P-PSDB; ABP73486.
 XX
 PT Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele of
 PT a gene and placing other allele of the gene under conditional expression.
 XX
 PS Claim 37; SEQ ID NO 6323; 167pp + Sequence Listing; English.
 XX
 CC The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office
 XX
 SQ Sequence 2256 BP; 648 A; 336 C; 456 G; 816 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,16e-272 Length: 2256
 Score: 2738.00 Matches: 509
 Percent Similarity: 81.22% Conservative: 110
 Best Local Similarity: 69.25% Mismatches: 110
 Query Match: 72.80% Indels: 28
 DB: 6 Gaps: 3

US-10-028-384-6 (1-718) x ABZ32036 (1-2256)

QY 9 LeuSerValPheGlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAla 28
 DB 52 GTTGAACGATTAGAGTATTATTAAAGCTATTATTATATATCATGATGAGGTGACGCT 111
 QY 29 IleSerSerArgLeuPheAlaValIleLysPheGluSerIleLeuHisGluPheAspPro 48
 DB 112 ATTTCTTCCTCGTTTATTTTCGGTATTCGATTGAAAGTATTATTATCATGATTCGATCCT 171
 QY 49 TrpPheAsnTyrArgAlaThrLysTyrLeuValAsnSerPheTyrLysPheLeuAsn 68
 DB 172 TGGTTCAATTTCCGAGCAACCAATATTATCTCACTCATTCCTTTTATGAATTTTGAAT 231
 QY 69 TrpPheAspArgThrTyrProLeuGlyArgValThrGlyGlyThrLeuTyrPro 88
 DB 232 TGGTTTGATGATAGAACTTGGTACCATTTGGAGAGCTCACTGGTGACTTTATATCCC 291
 QY 89 GlyLeuMetThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuPro 108
 DB 292 GGTTTAATGGTGACTTCAGTCCCATTTGGCATATTTTACGTGATTTGGTTGGCTTACCC 351
 QY 109 IleAspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrp 128
 DB 352 GTTGATATTAGAAATATTGTGTTTATTAGCACCACTTTCTCGGATTAACGTCAATT 411
 QY 129 AlaThrTyrGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaGly 148
 DB 412 TGTACTTATTTTTCGACTAAAGAAATGAAGGATTTCTAGTCAGGATTTATGGCAGCTATA 471
 QY 149 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168
 DB 472 TTATGGGATTTGCCCGAGGTATATTTCAAGATCAGTGGCTGTTCTTATGATTAATGAA 531
 QY 169 AlaIleAlaIleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThr 189
 DB 532 GCRAATTCATTTACTTTTAAATGCAACATTTTATTTCTGGATTAATCAATGAAATG 591
 QY 189 GlySerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrp 208
 DB 592 GGTTCAGTTTCTATGCCACATTTGACAGCATTTATTTCTATATGTTAGTGTCTGG 651
 QY 209 GlyGlyTyrValPheIleThrAsnLeuLeuProLeuHisValPheLeuLeuLeuMet 228
 DB 652 GGTGGATATGTTTTCATTTACCAATTTGATTCATTTACACGTAATTTGCTTGAATTTTCATG 711
 QY 229 GlyArgTyrSerSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrVal 248
 DB 712 GGTGCTTATTAATGCCAACTTTACACTGCTTATACATGATGATGCTTGGTACTTTG 771
 QY 249 AlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAla 268
 DB 772 GCATCAATGCAGATTCCTTCCTGGGTTTTTACCAATTAGATCAATGATCATATCGCT 831
 QY 269 AlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGln 288
 DB 832 GCATTAGAGTATTGGATTGTTACAAATTAGTGGCTTTGGTATTATGTTAATCAAAA 891
 QY 289 IleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeu 308
 DB 892 GTTCCAAACAAACAAATTAATCATTTCTGTAGTATTCATTTGCTGTGTGTGATTA 951
 QY 309 GlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuLeuAlaProTyrThrGlyArg 328

Db 952 GGTATTGGTGGATTATTGCGAATAACAGCAATGGGTTGGATTGCTCTTGGACAGGTAGA 1011
 Qy 329 pheTySerLeuThrAspThrAsnTyAlaLysIleHisIleProIleAlaSerVal 348
 Db 1012 TTTTATTCTCTATGGGATACAAATATATGCAAGATTCATATCCAAATATTGCTTCGTT 1071
 Qy 349 SerGluHisIleProValSerProAlaPhePheAspThrHisPheLeuIleTyr 368
 Db 1072 TCGAATCAACCACTACTGCTGGCCAGCAATCTTTTCGATCTAGTAGCTTATTTGG 1131
 Qy 369 LeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheVal 388
 Db 1132 TTATTCGCCGCTGCTATCTATTATGTTTCCAGAAATGCAAGGATGAACAGCTTTTCATT 1191
 Qy 389 IleAlaTySerValLeuTySerTyPheAlaGlyValMetValArgLeuMetLeuThr 408
 Db 1192 ATCATTTACAGTATGTTCTTATTGCTGGTGTCATGCTAAGATTGAATGTTGACT 1251
 Qy 409 LeuThrProValIleCysValSerAlaValAlaLeuSerLysIlePheAspIleTyr 428
 Db 1252 TTGACTCCAGTCATTTGCTGTCAGCAATTCGCTTATCTAAATGTTTGTGATGCTAT 1311
 Qy 429 LeuAspPhe----- 431
 Db 1312 TTGGACATTTGCTTCTACTCGAAGGTTGGAAAGTATGATGACGTTAGTGAC 1371
 Qy 432 -----LysThrSerAspArgLysTyAlaLysProAlaAlaLeu 445
 Db 1372 GAATCCAGAGCTCAACCAAAATCAAGTTCAGATTTCCAAAT-----GCTGGATAT 1425
 Qy 446 LeuAlaLysLeuIleValSerGlySerPheIlePheTyLeuTyLeuPheValPheHis 465
 Db 1426 TTGCTCAAAAGTTTGGTTTACTGACATTTACATTTTACCTTTCTACTTGTGTTTACAT 1485
 Qy 466 SerThrTyrValThrArgThrAlaTySerSerProSerValValLeuProSerGlnThr 485
 Db 1486 TGACTTGGGTAACTCGAATGCTTATTCATCACCATCAGTGTGTTTAGCATCCAGAAAC 1545
 Qy 486 ProAspGlyLysLeuAlaLeuIleAspPheArgGluAlaTyTyrTyrTyrLeuArgMet 505
 Db 1546 CCAGATGGCTCACACATATCATTTGATGATTATGAGAGGCTTATCTGTTAAGATG 1605
 Qy 506 AsnSerAspGluAspSerLysValAlaAlaTyrTrpAspTyGlyTyrGlnIleGlyGly 525
 Db 1606 AATACACCAAGATCCCAAGTATGCGCTGGTGGATTATGGTTATCAAAATCGGGGT 1665
 Qy 526 MetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIle 545
 Db 1666 ATGGCTGATAGAACACACACTTGTGTGATAACAATACATGGAATAACACACATATTGCCACT 1725
 Qy 546 ValGlyLysAlaMetAlaSerProGluGluLysSerTyGluIleLeuLysGluHisAsp 565
 Db 1726 GTTGGTAAGCAATGCTTCCCTGAAGATGTGCTATGAATTTTGAGACACACGAT 1785
 Qy 566 ValAspTyValLeuValIlePheGlyGlyLeuIleGlyPheGlyGlyAspAspIleAsn 585
 Db 1786 GTTGATTATGTAGTATTATTTGAGGGTTATTGGGTTATTCTGCTGGATGATATTAAAC 1845
 Qy 586 LysPheLeuTrpMetIleArgIleSerGluGlyIleTyrProGluGluIleLysGluArg 605
 Db 1846 AAATCTTATGATGTAGATGATGCTGGAAGTATCTGCTGATGAATATCAAGAAAGA 1905
 Qy 606 TyrPheTyThrAlaGluGlyGluTyArgValAlaAspAlaArgAlaSerGluThrMetArg 625
 Db 1906 GACTATTCTACTGCCGAGGAGAAATAAAGTGGATAAAGATGCATCATCTGCCAATGAAG 1965
 Qy 626 AsnSerLeuTyLysMetSerTyLysAspPheProGlnLeuPheAsnGlyGlyGln 645
 Db 1966 AATCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2025
 Qy 646 AlaThrAspArgValArgGlnGlnMetIleThrProLeuAspValProProLeuAspTy 665
 Db 2026 GGTGTTGATAGATTAGAAACCAACCAAAATCCCGCAAGATGAAGTACCCGAAATTTGAATGTT 2085

Qy 666 PheAspGluValPheThrSerGluAsnTrpMetValArgIleTyGlnLeuLysLysAsp 685
 Db 2086 GTTGAAGAAGCCTTCACATCAGAAATTTGATTTGAGATTTTACAAAGTTAAGATTG 2145
 Qy 686 AspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArgSerSer----- 702
 Db 2146 GATTAATGTTGGTAGAGATTTACATCAAGCTACTGCTTTTGAAGATTCATCATCCGCACT 2205
 Qy 703 ThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArg 717
 Db 2206 TCCAAAGAAACAGATCCATTAAGAGACCTAAATTCGAAGTAGA 2250
 RESULT 5
 ID ABT20789 standard; DNA; 2232 BP.
 XX ABT20789;
 XX 16-APR-2003 (first entry)
 XX Aspergillus fumigatus essential gene #3147.
 XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 XX cancer; contamination; biofilm; antibody; immune response; ds.
 XX Aspergillus fumigatus.
 XX WO200286090-A2.
 XX 31-OCT-2002.
 XX 23-APR-2002; 2002WO-US013142.
 XX 23-APR-2001; 2001US-0285697P.
 XX 27-APR-2001; 2001US-0287066P.
 XX 05-JUN-2001; 2001US-0295890P.
 XX 09-JUL-2001; 2001US-0303899P.
 XX 31-AUG-2001; 2001US-0316362P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.

New purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*, useful for treating or preventing infections by *A. fumigatus*, or for treating a non-infectious disease in a subject e.g. cancer.

Disclosure; Page: 175pp; English.

The invention relates to novel purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as *A. fumigatus*, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or contain contamination of an object by *A. fumigatus*, or to prevent or inhibit formation on a surface of a biofilm comprising *A. fumigatus*. The polynucleotides are useful for expressing recombinant protein for characterisation, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of *A. fumigatus* to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparisons with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively

[REDACTED]

22

635 LysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMet 654
 1978 TACAATTTCACACTCTCTTCCGTCGGGCCAAGCTGTCACCGGTCGGTGCCTCAAAA 2037
 655 IleThrProLeuAspValProProLeuAspTyrPheAspGluValPheThrSerGluAsn 674
 2038 CTT---CCACAGAGAGCCCTCAGCTCTCTACACTCGAAGAAGCTTTTCAGAGCGAGAAC 2094
 675 TrpMetValArgIleTyrGlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAsp 694
 2095 TGGATCATTCGTATCTACAGCTCAAGGATCTTGACAACTTGGCCGAGCACCACACAG 2154
 695 ValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLysArg 711
 2155 GCTGTTGCTTCGACAAA---GGTCTCAAGAAAAGCCGAGTACAAAAGAGG 2202

RESULT 6

ADB69900
 ID ADB69900 standard; DNA; 2466 BP.

AC ADB69900;

XX 04-DEC-2003 (first entry)

XX C. neoformans open reading frame SEQ ID NO:2305.

DE da; gene; fungicide; gene therapy; infection.

XX Cryptococcus neoformans.

XX WO2003052076-A2.

XX 26-JUN-2003.

XX 17-DEC-2002; 2002WO-US040225.

XX 17-DEC-2001; 2001US-0341261P.

XX (ELIT-) ELITRA PHARM INC.

XX Zamudio C, Eroshkin AM;

XX WPI; 2003-533017/50.

XX P-PSDB; ADB70261.

XX New nucleic acid, useful for preparing a composition for treating an

XX infection caused by Cryptococcus neoformans.

XX Claim 2; SEQ ID NO 2305; 136pp; English.

XX The invention relates to a novel purified or isolated Cryptococcus
 CC neoformans nucleic acid molecule comprising a sequence encoding a
 CC polypeptide comprising a sequence not given in the specification. A
 CC polynucleotide of the invention has nuclease activity, and may have a
 CC use in gene therapy. The nucleic acid is useful for preparing a
 CC composition for treating an infection caused by Cryptococcus neoformans.
 CC The present sequence represents a C. neoformans sequence of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 2466 BP; 508 A; 678 C; 580 G; 700 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.99e-227	Length:	2466
Score:	2301.50	Matches:	435
Percent Similarity:	73.05%	Conservative:	107
Best Local Similarity:	58.63%	Mismatches:	155
Query Match:	61.13%	Indels:	45
DB:	10	Gaps:	6

US-10-028-384-6 (1-718) x ADB69900 (1-2466)

QY 13 GlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArg 32
 DB 223 GAACGCTTTGCGCTTCATCATCTCGCGGTTAATATGCGGTCTGCGATGGAAGTCGA 282
 QY 33 LeuPheAlaValIleLysPheGluSerIleIleHISqluPheAspProTrpPheAsnTyr 52
 DB 283 TTGTTTGGCGTGATCAGATTGCAATCTGTCTATCCACGAATTCACCCCTGGTCAACTAC 342
 QY 53 ArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnTrpPheAspAsp 72
 DB 343 CGAGCTCGAAGCTCTTGTTTACAGGGTTCTACGAGTTCTTGGAACTGCTTTGACCCC 402
 QY 73 ArgThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThr 92
 DB 403 TCCGCTTGCTACCTCTCGGCAGAACTGCTGCTACCAAGCTCTATCTCGCTTGGTGC 462
 QY 93 ThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArg 112
 DB 463 ACGTCTGGACTGATTTGGCATGCTCTTCGGGCA---ATCAATATGCCGCGGACATTCGC 519
 QY 113 AsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGlu 132
 DB 520 AATGTCTGTCTCTTGCACCTGCGATTTTCTGGATTGACTGCTGCGGCGACTTATCTT 579
 QY 133 PheThrLysGluIleLysAspAlaSerAlaGlyLeuAlaAlaGlyPheIleAlaIle 152
 DB 580 TTCACCACTGAATGCTACACCATCAGCTGCTTATTCGCGCGCGCTTCATGGCATT 639
 QY 153 ValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIle 172
 DB 640 GTACCGGATACATCTCTCGATCTGTCGCGGTTCTTATGACCAACGAAGCATTGCGCATC 699
 QY 173 ThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMet 192
 DB 700 TTCCTCTGTGAGCTCTCTACTCTTGGATTAGCCCGTCAAAACCGTAGCTCATTT 759
 QY 193 HisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrVal 212
 DB 760 TGGGATGATGATCACTGCTCTGTTCTACGGGTGGATGCTTCTCATCGGCGGTGTTACGTT 819
 QY 213 PheIleThrAsnLeuLeuProLeuHisValPheLeuLeuLeuLeuMetGlyArgTyrSer 232
 DB 820 TTCAATCACCACATGATTCATTCGACGCTTGTCTCATTTGTCATGTCATGCGAGTCAAC 879
 QY 233 SerLysLeuTyrSerAlaTyrThrTyrTyrAlaIleGlyThrValAlaSerMetGln 252
 DB 880 AACCGGCTTTATACCGCTTACTCTCTCTGTTATGTCATTGGAACATATCGCTCCATGAG 939
 QY 253 IleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGlyVal 272
 DB 940 GTCCCTTTGGAGTTCTCTCCCATCCGAACCTCTGAGCAGCATGCGGCGCTTGGTGT 999
 QY 273 PheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAla 292
 DB 1000 TTCGCTCTTGACAGCTGATCGGATTCGTGAAAGTCTGTCGACGACTCGTCTGCGTGGCAAG 1059
 QY 293 LysPheLysValIleMetMetValSerLeuPheLeuLeuLeuValLeuGlyValGly 312
 DB 1060 CAATTCAGCTCTCTCAAGCTTTCGTCGCGCGTATTCGCTCAGTTTGTCTGCGC 1119
 QY 313 LeuSerAlaLeuThrTyrMetGlyLeuLeuAlaProTrpThrGlyArgPheTyrSerLeu 332
 DB 1120 CTCGTCATTTGACTTTCTCTGATGATGATGCGCCCTCTGCGTGAAGATTTTATCTCTT 1179
 QY 333 TrpAspThrAsnTyrAlaLysIleHisIleProIleAlaSerValSerGluHisGln 352
 DB 1180 TGGGATAGTGGCTATGGAAGTCCACATGCCCATTTATTCCTCTCCGACACACAG 1239
 QY 353 ProValSerTrpProAlaPhePheAspThrHisPheLeuLeuTrpPheProAla 372
 DB 1240 CCCACCGCTTGGCGCTCATTTCTACTTGTACCTCGAAGTGTATCTCTTTTCTGCTGCC 1299
 QY 373 GlyValPheLeuLeuPheLeuAspGluHisValPheValIleAlaTyrSer 392

DB	2377	GTACTGCTCGGACGCGGGTGAAGTGTGAAGAGAGAGTCTTAGTCCAGTAGGCGGTG	2436
QV	710	LysArg 711	
DB	2437	AAGCGG 2442	
RESULT 7			
ABT20191			
ID	ABT20191	standard; DNA; 2603 BP.	
XX	AC		
XX	ABT20191;		
XX	16-APR-2003	(first entry)	
DT			
DE	Aspergillus fumigatus	essential gene #2549.	
XX			
XX	Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;		
XX	cancer; contamination; biofilm; antibody; immune response; ds.		
XX			
OS	Aspergillus fumigatus.		
XX			
XX	WO200286090-A2.		
XX			
PD	31-OCT-2002.		
XX			
PF	23-APR-2002; 2002WO-US013142.		
XX			
PR	23-APR-2001; 2001US-0285697P.		
XX	27-APR-2001; 2001US-0287066P.		
PR	05-JUN-2001; 2001US-0295890P.		
PR	09-JUL-2001; 2001US-0303899P.		
PR	31-AUG-2001; 2001US-0316362P.		
XX			
XX	(ELIT-) ELITRA PHARM INC.		
XX			
PI	Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;		
XX			
XX	WPI; 2003-093124/08.		
DR			
PT	New purified or isolated nucleic acids of essential genes of Aspergillus		
PT	fumigatus, useful for treating or preventing infections by A. fumigatus,		
PT	or for treating a non-infectious disease in a subject e.g. cancer.		
PS	Disclosure; Page: 175pp; English.		
XX			
XX	The invention relates to novel purified or isolated nucleic acids of		
CC	essential genes of Aspergillus fumigatus. The isolated nucleic acids of		
CC	the invention are used to treat or prevent infections by a pathogenic		
CC	organism such as A. fumigatus, to treat a non-infectious disease in a		
CC	subject (e.g. cancer), to prevent or contain contamination of an object		
CC	by A. fumigatus, or to prevent or inhibit formation on a surface of a		
CC	biofilm comprising A. fumigatus. The polynucleotides are useful for		
CC	expressing recombinant protein for characterization, screening or		
CC	therapeutic use, as markers for host tissues in which the pathogenic		
CC	organisms invade or reside, for comparing with the DNA sequence of A.		
CC	fumigatus to identify duplicated genes or paralogues having the same or		
CC	similar biochemical activity and/or function, for comparing with DNA		
CC	sequences of other related or distant pathogenic organisms to identify		
CC	potential orthologous essential or virulence genes, for selecting and		
CC	making oligomers for attachment to a nucleic acid array for examination		
CC	of expression patterns, for raising anti-protein antibodies, as an		
CC	antigen to raise anti-DNA antibodies or to elicit another immune		
CC	response, and for identifying polynucleotides encoding the other protein		
CC	with which binding occurs or to identify inhibitors of the binding		
CC	interaction. The polypeptides may be used to raise antibodies or to		
CC	elicit immune response, as a reagent in assays designed to quantitatively		
CC	determine levels of the protein in biological fluids, as a marker for		
CC	host tissues in which pathogenic organism invade or reside, and to		
CC	isolate correlative receptors or ligands in the case of virulence		
CC	factors. This polynucleotide sequence represents one of the essential		
CC	genes of Aspergillus fumigatus of the invention		
XX			
XX	Sequence 2603 BP; 565 A; 698 C; 620 G; 720 T; 0 U; 0 Other;		

QY 572 lePheGlyGlyLeuIleGlyPheGlyGlyAspPheLeuLeuMetIleA 592
 DB 2097 TGTTCGGTGGTCTAGTATTCCTGGCATACATTAACAAATCTATGATGGTCC 2156
 QY 592 gIleSerGluGlyIleTrpProGluGluIleGlyGlyGlyGlyGlyGlyGly 612
 DB 2157 GTATCGCGAAGGTATCTGCGCGATGAGTTAAAGAGCGGACATCTTTACTGCACCG 2216
 QY 612 lYGluTyArgValAspAlaArgAlaSerGluThrMetArgSerLeuLeu----- 629
 DB 2217 GTGAATATCGTGTGACGATGGAGCGACCCCACTATGCGACAGCTTGATGGAT 2276
 QY 630 -----Tyr 630
 DB 2277 TCCCTTCCTCCCTCTTATCTACTGCGTGCATCAGCTAATACAAATCTAGGTAT 2336
 QY 631 LysMetSerTyLysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
 DB 2337 AAAATGCTTATTACAAATTTCAACTCTCTCTCCGTCGGGCAAGCTGCGACCGGTC 2396
 QY 651 ArgGlnMetIleThrProLeuAspValProLeuAspTyPheAspGluValPhe 670
 DB 2397 CGTGGTCAAACTT---CCACAGAGAGCGCTCTCTACTCTCACTCGAAGAGCTTTC 2453
 QY 671 ThrSerGluAsnTrpMetValArgIleTyGlnLeuLysLysAspAlaGlnGlyArg 690
 DB 2454 ACGAGCGAGAACTGATCATCTGATCTATCAAGGTCAAGGTCTTGACAACTTGGCGGA 2513
 QY 691 ThrLeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgSerIleLys 710
 DB 2514 GACCACACACAGGTGTGTGCTTCGACAAA---GGTCTCAAGAAAAGCGGAGTCAAG 2570

RESULT 8
 ABT17781
 ID ABT17781 standard; DNA; 3969 BP.
 AC ABT17781;
 DT 16-APR-2003 (first entry)
 DE Aspergillus fumigatus essential gene #139.
 KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 OS Aspergillus fumigatus.
 PN WO200286090-A2.
 PD 31-OCT-2002.
 PF 23-APR-2002; 2002MO-US013142.
 PR 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Jiang B, Tishkoff D, Zamudio C, Broshkin AM, Hu W, Lemieux SM;
 WPI; 2003-093124/08.
 XX New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.

XX
 PS
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterization, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune responses, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organisms invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX
 SQ Sequence 3969 BP; 964 A; 996 C; 953 G; 1056 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,7e-223 Length: 3969
 Score: 2264.50 Matches: 461
 Percent Similarity: 66.31% Conservative: 98
 Best Local Similarity: 54.69% Mismatches: 135
 Query Match: 60.21% Indels: 150
 DB: 8 Gaps: 11
 US-10-028-384-6 (1-718) x ABT17781 (1-3969)
 QY 15 IleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPhe 34
 DB 698 CTTCTCGGATATTATCTCTGTGATACATTCCAGCAGCAGCAGTGCAGTGCAGCTTTC 757
 QY 35 AlaValIleLys----- 38
 DB 758 AGCGTTATCCG-TAAGTCATCGAGGAATCAAGTTGCTAAATTTGTTAATCTATCATCA 816
 QY 39 -----PheGluSerIleIleHisGlu----- 45
 DB 817 AATTGACTGATCATTCTTTCTTACCAGGCTTCGAGAGTATCATCCAGCAATGTAAGTAT 876
 QY 45 ----- 45
 DB 877 AGTCAATTTGTATACCTACCGGCTTCGGCCCTTTAACGGACCTCGCTTGGACGAATTGATA 936
 QY 46 -----PheAspProTrpPheAsnTyArgAlaThrLysTyLeuVal 59
 DB 937 CTGATGCTCCGATATAGTTGACCCGTTGTTCAACTTCGAGAGCAACAAATACTTAGTA 996
 QY 60 AsnAsnSerPheTyLysPheLeuAsnTrpPheAsp----- 72
 DB 997 CAGATGGTTTCTATAGCTTTTGGGATTTGGTTGATGACCGTATGTGGTCCCGACCGCA 1056
 QY 72 ----- 72
 DB 1057 CAAGAGTGCATGAGCAACAGGTACCTGTCTCGAGTGTCTTAGGCTAATCGGGTCC 1116
 QY 73 -----Arg-ThrTrpTyProLeuGlyArgValThrGlyGlyThrLeuTyProGlyLe 90
 DB 1117 CAACATAGGAACATGGCATCTCTTGGGACAGTGTCCCGGTGGACGATTATATCCCGTCT 1176

```

Qy 90 uMetThrThrSerAlaPheIleThrHisAlaLeuArgAsnTrpLeuGlyLeuProIleAs 110
Db 1177 CATGGTACGAGCGCGGTGATACCATATCTTTGCGA---TTCTTACTACTCCCGTCGA 1233
Qy 110 PileArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaTh 130
Db 1234 TATTGCAACATCTCGCTCTACTGCGCCAGGATTCCTCGGCTGACTGCAATGGAAT 1293
Qy 130 rTyCluPheThrIysGluIleLysAspAla---SerAlaGlyLeuLeuAlaGlyPh 149
Db 1294 GTACTGTGTCATCCGAGATGCTCTCTCGCATCTGCAAGGCTTCTTTCGACAGCTTT 1353
Qy 149 eileAlaIleValProGlyTrpIleSerArgSerValAlaGlySerTyAspAsnGluAl 169
Db 1354 CATGGGAATCGCCCTGGTACATCTCCGATCAGTTGCTGGAAGCTACGATAAGCAAGC 1413
Qy 169 aileAlaIleThrLeuLeuMetValThrPheMetPheTrpIleValAlaGlyThrCl 189
Db 1414 GATTGCCAATCTTCTGCTTGTGTCATCTTCTTCTGATCAAGGCTGCAAAATGG 1473
Qy 189 ySerIleMetHisAlaThrCysAlaAlaLeuPheTyThrPheTyMetValSerAlaTrpGl 209
Db 1474 GTCTATCATGTGGGAGCGGTGACCGCACATATTCTACGGCTACATGGTGTGCGCATGGG 1533
Qy 209 yGlyTyValPheIleThrLeuIleProLeuHisValPheLeuLeuIleLeuMetCl 229
Db 1534 TGGGTATGCTCTCATACGACCTGATCCCGCTGACGATTTTGTCTTCTGTGCGATGG 1593
Qy 229 yArgTySerSerIysLeuTySerAlaTyThrTrpTyThrAlaIleGlyThrValAl 249
Db 1594 TAGATACAGCACTCGCATCTACATTAGCTATACCATGATGATGCGTGGGACTTTGGC 1653
Qy 249 aSerMetClnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAl 269
Db 1654 TAGATGAGATTCCTCTGCGGATTTTGGCTATCCGAACAGCGGCACATGTCGCG 1713
Qy 269 aLeu-----G 271
Db 1714 CTGGGTAGGTCCTTTCGCAAGTCAATGGCGGCTTCAGGAGTCGACATTTTGCAG 1773
Qy 271 lyValPheGlyLeuIleClnIleValAlaPheGlyAspPheValLysGlyIleSerT 291
Db 1774 GTGCTTCGGCTGCTCAGCTTGTGGCTTGCCTGCGAGTTTGTCCGAGCTTCGTTCCAA 1833
Qy 291 hrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeuGlyValV 311
Db 1834 GCAAGCAGTTCAGAGACTTCTGACCGCATGATCTTCATCACCTTCGGTCTCGGTTTC 1893
Qy 311 alGlyLeuSerAlaLeuTyThrMetGlyLeuIleAlaProTyThrGlyArgPheTyS 331
Db 1894 TTGGCTAGTGTCTGACTGTGACGGGAGTGATCGCTCTTGGAGCGCGGATTTCTACT 1953
Qy 331 erLeuTrpAspThrAsnTyAlaLysIleHisIleProIleAlaSerValSerGluH 351
Db 1954 CTTTGTGGGACACTGGCTATGCCAAATCCATCTCCATCATCTGCTCAGTCTCGGAAC 2013
Qy 351 isGlnProValSerTrpProAlaPhePhePheThrHisPheLeuIleTrpLeuPheP 371
Db 2014 ACCAGCCACCGCTTGGCAGCGTCTCTTCGATCTGAATCTTCTGATCTGCTGCTTTCC 2073
Qy 371 roAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAla 391
Db 2074 CGGCAGGTGTCTACATGTCTTCGCTGACCTCAAGGACGAGCATGCTTCGTCATTATCT 2133
Qy 391 yrSerValLeuCysSerTyThrPheAlaGlyValMetValArgLeuMetLeuThrP 411
Db 2134 ACTCGCTCTTCGAGCTACTTCGCCGGTGATGATGGTCCGACATATGCTGACCTTGACC 2193
Qy 411 roValIleCysValSerAlaAlaValAlaLeuSerIysIlePheAspIleTyLeu---- 429
Db 2194 CTATTGTGTGTGGCGCTCTCTGCGCTGTGCTCCATCTTCGACACCTATATGGCGA 2253

```

RESULT 9
 ABT19595
 ID ABT19595 standard; DNA; 4603 BP.
 XX
 AC ABT19595;

```

Qy 430 -----AspPheIysThrSerAsp----- 435
Db 2254 CTACCTCCCGACACACGAGCTGTGAAGCGAAACGAATAGAGACTCGTCTTCACAACTC 2313
Qy 436 -----ArgLysTyTrAlaIleLysProAlaLeuLeuAlaLysLeuIleValS 452
Db 2314 TTCGCTCAGTTCGGAAGCCCAATGTGGAAATCACCTCCCATGTTTCTAAGATTATAGTGA 2373
Qy 452 erGlySerPheIlePheTyLeuPheValPheHisSerThrTrpValThrArgT 472
Db 2374 CGCGCTGTGTGCTGCTACCTGCTCTCTGTTTGTGGCAGCTGCACCTGGGTACTACGA 2433
Qy 472 hrAlaTySerSerProSerValValLeuProSerClnThrProAspGlyLysLeuAla 492
Db 2434 ATGCATACTCTTCTCTCTCGTGTCTTGGCTAGTCCGATGCTCAGCGAAGCCCAATACA 2493
Qy 492 euIleAspAspPheArgGluAlaTyTrpLeuArgMetAsnSerAspGluAspSerL 512
Db 2494 TCATTGACGATTAATCGTGGCTTACTACTGCTTCGTCAGATTAATCTCTCAGAACGCCA 2553
Qy 512 ysValAlaAlaIleTrpTyrGlyTyGlnIleGlyMetAlaAspArgThrL 532
Db 2554 AAATCATGTCATGCTGGATTAATGGTATCAAAATCGTGGCATGGCGACCCCAACT 2613
Qy 532 euValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAla 552
Db 2614 TGGTTGACAAACAACCTGGAACACACCCATTAATGCTACGTTGGTAAGGAGTAGCT 2673
Qy 552 erProGluGluLysSerTyGluIleLeuLysGluHisAspValAspTyValLeuValI 572
Db 2674 CACGCGAGGAAAGTCAGCTACCCCATCTCCGCGCAGCATGATGCTGATTCGTCGTGG 2733
Qy 572 lePheGlyGlyLeuIleGlyPheGlyGlyAspIleAsnLysPheLeuTrpMetIleA 592
Db 2734 TGTTCGCTGCTGCTAGTATTAATCTGGCGATGACATTAACAAATCTTATGATGGTCC 2793
Qy 592 rgIleSerGluGlyIleTrpProGluGluIleLysGluArgTyPheTyThrAlaGluG 612
Db 2794 GTATCGCGAAGGTATCTGCGCGATGAGGTAAAGAGCGGACTTCTTTACTGCACGCG 2853
Qy 612 lyGluTyArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeu----- 629
Db 2854 GTGATATCTGTGCGAGTGGAGCGACCCCACTATGGCGAACAGCTTGATGATGAT 2913
Qy 630 -----Tyr 630
Db 2914 TCCTCTCTCTCTCTTATCTACTGCTGCTGATCAACGCTAAATTCACAAATCTAGGTAT 2973
Qy 631 LysMetSerTyTrLysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
Db 2974 AAATGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3033
Qy 651 ArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyPheAspGluValPhe 670
Db 3034 CGTGGGTCAAAACTT---CCACAGAAAGGCCCTCAGCTCTCTACACTCAGAGAGCTTTC 3090
Qy 671 ThrSerGluAsnTrpMetValArgIleTyGlnLeuLysLysAspAlaGlnGlyArg 690
Db 3091 ACAGCGAGAGACTGATCATCTGATTAATCAAGTCAAGGATCTTCAGAACCTTGGCGCA 3150
Qy 691 ThrLeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLys 710
Db 3151 GACCACACACCGAGCTGTGCTTTCGACAAA---GGTCTCAGAAAAAGCGGAGTACAAG 3207
Qy 711 Arg 711
Db 3208 AGG 3210

```


Db 2137 GTGTCCTCGCCTGCTAGCTTTGGCCCTTGGCGAGTTTGTCCGAGCTTGCTTCCAA 2196
QY hrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeuGlyValV 311
Db 2197 GCAAGCAGTTCCAGAGACTTCTGACCGCATGATCTTCATCACCTTCGGTCTCGTTTC 2256
QY 311 aGlyLeuSerAlaLeuThrTyMetGlyLeuLeuAlaProThrGlyArgPheTyrs 331
Db 2257 TTGGGTAGTTGTTGATGCTGACGGAGGATGCTCTCTGAGGCGCGGATCTTACT 2316
QY 331 erLeuTrpAspThrAsnTyAlaLysIleHisIleProIleAlaSerValSerGluH 351
Db 2317 CTTTGGGACACTGGCTATGCAAAATCCACATTCCTCATTCCTGCTGCTCGGAAC 2376
QY 351 iGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPheP 371
Db 2377 ACCAGCCCACTGGCCGAGTCTTCTTCGATCTGAATCTCTGATCTGGCTTTTC 2436
QY 371 roAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAla 391
Db 2437 CGGAGGTCTACATGCTTCCGTGACCTCAAGGACGAGCATGCTTGTGCTATTATCT 2496
QY 391 yrSerValLeuCysserTyPheAlaGlyValMetValArgLeuMetLeuThrP 411
Db 2497 ACTCGTCTCGGAGCTACTTGGCCGGTGTATGGTCCGACTAATGCTGACCTTGACC 2556
QY 411 roValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyLeu 429
Db 2557 CTATTGTGTGTGCGCTGCTCTGCGCTGTCTGCTCATCTCGACACCTATATGCGGA 2616
QY 430 -----AspPheLysThrSerAsp----- 435
Db 2617 CTACCTCCGACACAGGCTGGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2676
QY 436 -----ArgLysTyAlaIleLysProAlaAlaLeuLeuAlaLysLeuVal 452
Db 2677 TTCGCTCAGTTCGGAAGCCCAATGTGGAATCACCTCCCATGTTCTTAAGATTATAGTA 2736
QY 452 erGlySerPheIlePheTyLeuTyLeuPheValPheHisSerThrTrpValThrArgT 472
Db 2737 CGCGTCTGTTGCTTACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2796
QY 472 hrAlaTySerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAla 492
Db 2797 ATGCATACCTCTCTCTCCGTTGCTTGGCTAGTGGCTGCTGCTGCTGCTGCTGCTGCT 2856
QY 492 euIleAspPheArgGluAlaTyTyTrpLeuValMetAsnSerAspGluAspSerL 512
Db 2857 TCATTGACATTATCGTAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2916
QY 512 yaValAlaAlaTrpTrpAspTyGlyTyGlnIleGlyMetAlaAspArgThrL 532
Db 2917 AAATCATGCTGCTGGGATGATGCTTCAATCGGTGCTGCTGCTGCTGCTGCTGCTGCT 2976
QY 532 euValAspAsnThrTrpAsnThrHisIleAlaIleValGlyValAlaMetAla 552
Db 2977 TGTGTTGACACACACCTCGAACACACCCATATGCTACCGTTGGTAAGGCGATGAGCT 3036
QY 552 erProGluGlySerTyGluLeuLeuLysGluHisAspValAspTyValLeuValI 572
Db 3037 CACGCGAGGAGTACCTACCCATCTCCGCGAGCATGATGCTGCTGCTGCTGCTGCTGCT 3096
QY 572 lePheGlyLeuIleGlyPheGlyAspAspIleAsnLysPheLeuTrpMetIleA 592
Db 3097 TGTGCGGTCTGCTAGTATTCTGCGGATGATTAACAAATCTTATGATGCTGCT 3156
QY 592 rgIleSerGlyIleTrpProGluGluIleLysGluArgTyPheTyThrAlaGluG 612
Db 3157 GTATCGCGAGGATATCTGCGCGGATGAGGTAAAGCGGAGCTTCTTACTGCGCG 3216
QY 612 lyGlyTyArgValAlaAlaAlaSerGluThrMetArgAsnSerLeuLeu----- 629

Db 3217 GTGAATATCGTGTGCAAGATGAGAGAGCCCAAACTATGCGCAACAGCTTGATGATGAT 3276
QY 630 -----Tyr 630
Db 3277 TCCCTCTCTCCCTCTTATCTACTGCGTGCATCAAGCTAATTCACAAAATCTAGGTAT 3336
QY 631 LysMetSerTyLysAspPheProGluLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
Db 3337 AAAATGTCTTATTAACAATTTCAATCTCTTCCGTCGGGCGCAAGCTGTCGACCGGTC 3396
QY 651 ArgGlnGlnMetIleThrProLeuAspValProLeuAspTyPheAspGluValPhe 670
Db 3397 CGTGGTCAAAACTT-----CCACAGAAGGCCCTCAGCTCTCTACCTCGAAGAAGCTTTC 3453
QY 671 ThrSerGluAsnTrpMetValArgIleTyGlnLeuLysLysAspAlaGlnGlyArg 690
Db 3454 ACGAGCGAAGACTGGATCAITTCATCAAGGTCAAGATCTTACAACTTGGCCGA 3513
QY 691 ThrLeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLys 710
Db 3514 GACCACACACAGGCTGTTGCCCTTCGACAAA---GGTCTCAAGAAAAGCGGAGTACAAAG 3570
QY 711 Arg 711
Db 3571 AGG 3573
RESULT 10
ABT18969
ID ABT18969 standard; DNA; 1848 BP.
XX AC ABT18969;
XX XX
DT 16-APR-2003 (first entry).
XX DE Aspergillus fumigatus essential gene #1327.
XX DE Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response; ds.
XX OS Aspergillus fumigatus.
XX XX
XX PN WO200286090-A2.
XX PD 31-OCT-2002.
XX PF 23-APR-2002; 2002WO-US013142.
XX PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287086P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX PS Disclosure; Page; 175pp; English.
XX CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or

CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogs having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention
 XX Sequence 1848 BP; 363 A; 510 C; 447 G; 528 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,01e-205 Length: 1848
 Score: 2090.00 Matches: 394
 Percent Similarity: 79.53% Conservative: 80
 Best Local Similarity: 66.11% Mismatches: 93
 Query Match: 55.57% Indels: 29
 DB: 8 Gaps: 5

US-10-028-384-6 (1-718) x ABT18969 (1-1848)

QY 63 PheTyrLysPheLeuAenTrrPheAspAsp-----Arg-ThrTr 75
 DB 6 TTCATAGCTTTGGGATTGGTTGATGACCGCTAATCGCGTCCCAACATAGGACATG 65
 QY 75 pTy-ProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrSerAl 95
 DB 66 GCATCCTCTGGGACGTGTCCCGGTGGCAGCTATATATCCCGTCTCATGTCGACGAGGG 125
 QY 95 aPheLeuTrpHisAlaLeuAenTrrPheLeuGlyLeuProLysPheAspLysValCy 115
 DB 126 CGTATCTACCATATCTTGGGA---TTCCTTATATCCCGTGGATTCGCAACATCTG 182
 QY 115 sValLeuPheAlaProLeuPheSerGlyValThrAlaTrrPheAlaThrTyrGluPheThrLy 135
 DB 183 CGTCTCTACTGGCGCCAGGATCTCCGCGCTGACTGTCATGGCAATGTACTTCGTGACATC 242
 QY 135 sGluLeuLysAspAla---SerAlaGlyLeuLeuAlaAlaGlyPheLeuAlaLeuValPr 154
 DB 243 CGAGATGCTCTCTCGCCATCTCGAGTCTCTTTGCGAGAGCTTTTCATGGGAATCGCCCC 302
 QY 154 oGlyTrrLysSerArgSerValAlaGlySerTyrAspAsnGluAlaLeuAlaLeuThrLe 174
 DB 303 TGGTTACATCTCCCGATCAGTTGGTGAAGCTACGATACGAAGCGATTGCCATCTCTCT 362
 QY 174 uLeuMetValThrPheMetPheTrrLysAlaGlnLysThrGlySerLysMetHisAl 194
 DB 363 GTTTGTGTTCACATCTTCTATGGATCAAGGCTGTCAAAATGGGTCTATCATGTGGGG 422
 QY 194 aThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrrPheGlyTyrValPheIl 214
 DB 423 AGCGCTGACCCCATATCTACGGCTACATGGTGTCCGATGGGGTGGTATGTTCTTCAT 482
 QY 214 eThrAsnLeuLeuProLeuHisValPheLeuLeuLeuLeuMetGlyArgTyrSerSery 234
 DB 483 TACGAACCTGATCCCTCGACGCTTTTGTCTTCTGTCATGGGTAGATACAGCACTCG 542
 QY 234 sLeuTyrSerAlaTyrThrTrrPheAlaLeuGlyThrValAlaSerMetGlnLePr 254
 DB 543 CATCTACATATGATACCATGGTATGGCGTGGGAGCTTTGGCTAGCATGCATGATCC 602
 QY 254 oPheValGlyPheLeuProLysArgSerAsnAspHisMetAlaAlaLeuGlyValPheGl 274

DB 603 CTTCTGCGGATTTTTCCTATCCGAAACAGCAGCACCATGTCCGCTTGGGTCTCTTCGG 662
 QY 274 yLeuLeuGlnLeuValAlaPheGlyAspPheValLysGlyGlnLysSerThrAlaLysPh 294
 DB 663 CTGTGTTAGTGTGGGCTTTTCGCGAGTTTGTCCGAGCTTGTTCGAAGCAAGCAGT 722
 QY 294 eLysValLeuMetMetValSerLeuPheLeuLeuValLeuValValGlyLeuSe 314
 DB 723 CCAGAGACTTCGACCGCCCATGATCTTCATCACCTCGGTCTGGTTGTGGGTAGT 782
 QY 314 rAlaLeuThrTyrMetGlyLeuLeuAlaProTrrPheGlyArgPheTrrSerLeuTrpAs 334
 DB 783 TGTTCGTGACTGACGGAGGATGCTCTTCCTTGAGCGCGCGATTCATCTCTCTGGGA 842
 QY 334 pThrAsnTyrAlaLysLysHisLeuProLysLeuAlaSerValSerGluHisGlnProVa 354
 DB 843 CACTGGCTATGCCAAATCCCATTCCTCATCTGCTCATCTGCGAACACACAGCCAC 902
 QY 354 lSerTrrProAlaPhePheAspThrHisPheLeuLeuTrrPheProAlaGlyVa 374
 DB 903 CGCTTGGCGAGGTTCTTCTTCGATCTGAACCTCTGATCTGGCTTTCCCGCAGGTGT 962
 QY 374 lPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValLeuAlaTrrSerValLe 394
 DB 963 CTACATGTGCTTCGCTGACCTCAAGGACGAGCATGTCTCTGTCATTATCTACTCGGTCT 1022
 QY 394 uCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValLeuCy 414
 DB 1023 TCGGAGCTACTTCGCGGTGTATGGTCCGACTAATGCTGACCTTGACCTTATTTGTG 1082
 QY 414 sValSerAlaAlaValAlaLeuSerLysLysPheAspLysTrrLeu----- 429
 DB 1083 TGTTCGCGTCTCTGCGCTGTCTGCTCATCTCGACACCTATATGCGGACTACCTCTCC 1142
 QY 430 -----AspPheLysThrSerAsp----- 435
 DB 1143 GACACAGCGCTCTGAAGCGAAAGCAAGATGAAGACTCGTCTTCAACCTCTTCGCTCAGT 1202
 QY 436 -ArgLysTyrAlaLeuLysProAlaLeuLeuAlaLysLeuLeuValSerGlySerPh 455
 DB 1203 TCGGAAGCCCAATGTGGAATCACCTCCCATGTTTCTAAGATTATAGTACGCGCTGT 1262
 QY 455 eLysPheTyrLeuTyrLeuPheValPheHisSerThrTrrPheValThrArgThrAlaTrrSe 475
 DB 1263 TGTGCTACCTGCTCTCTGTTGTCGCACTGCACTGGGTACATCGAATGCATCTC 1322
 QY 475 rSerProSerValLeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuAspAs 495
 DB 1323 TTCTCTTCCGTGTCTGGCTAGTGGATCGCTGACGGAAGCCAAATACATCATTCACGA 1382
 QY 495 pPheArgGluAlaTrrTyrTrrPheLeuArgMetAsnSerAspGluAspSerLysValAlaAl 515
 DB 1383 TTATCGTGGGCTTACTACTGCTTCGCAAGTACTCTCGAAGCCCAAAATCATGTC 1442
 QY 515 aTrrPheAspTyrGlyTrrGlnLeuGlyMetAlaAspArgThrThrLeuValAspAs 535
 DB 1443 ATGGTGGGATTATGGGTATCAATCGGTGGCATCGCGACCGCCCAACCTTGGTTCACAA 1502
 QY 535 nAsnThrTrrPheAsnThrHisLeuAlaLeuValGlyValAlaMetAlaSerProGluGl 555
 DB 1503 CAACACCTGGAAACACACCATATTTGCTACGTTGGTGAAGCGATGAGTCAACGCGAGA 1562
 QY 555 uLysSerTyrGluLeuLeuLysGluHisAspValAspTrrValLeuValLeuPheGlyGl 575
 DB 1563 AGTCAGCTACCCCATCTCCGCGACATGATGTCGATACGCTGCTGGTGTGGTGGTGG 1622
 QY 575 yLeuLeuGlyPheGlyGlyAspAspLysAsnLysPheLeuTrrMetLysArgLysSerGl 595
 DB 1623 TCTGCTAGGTATCTGGCATGACATTTACAAATCTTATGATGATGCTGCTGATCGCGGA 1682
 QY 595 uGlyLysLeuProGluGluLeuLysGluArgTyrPheTrrThrAlaGluGlyGluTrrAr 615
 DB 1683 AGGTATCTGGCCCGATGAGTTAAAGAGCGGAGCTTCTTACTGCGCGGTGAATATCG 1742

QY 334 spThrAsnTyrAlaValHisIleProIleAlaSerValSerGluHisGlnProv 354
 Db 963 ACACCTGGCTATGCCAAATACCATTCCTCATCTCCCTCAGTCTCGGAACACACGCCCA 1022
 QY 354 alSerTrpProAlaPhePheAspThrHisPheLeuLeuTrpPheProAlaGlyV 374
 Db 1023 CCGTTGGCCAGCGCTTCTTTCGATCTGAATCTCTGATCTGGCTTTCCCGGAGGTG 1082
 QY 374 alPheLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyrSerValL 394
 Db 1083 TCTACATGCTTCCGAGCTCAAGCAGCAGCATGCTTCGTCAATATCTACTCGGTCC 1142
 QY 394 euCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrProValIleC 414
 Db 1143 TTGGAGCTACTTCCCGCGGTGTTATGTCGATTAATGCTGACCTTGACCCCTATTGTGT 1202
 QY 414 ysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu 429
 Db 1203 GTGTTCGCGTCTCTGCGCTGTGCTGCATCTCTGACACCTATATGGGAGCTACCTCC 1262
 QY 430 -----AspPheIleThrSerAsp 435
 Db 1263 CGACACAGCGCTCTGAAGCGAAACGATGAAGACTGCTCTCACACCTCTTCCTCAG 1322
 QY 436 --ArgLysTyrAlaIleGlyProAlaAlaLeuLeuAlaLysLeuValSerGlySerP 455
 Db 1323 TTCGGAAGCCCAATGTTGGAATCACCTCCCATGTTCTAAGATTATAGTACGCGGTCTG 1382
 QY 455 heIlePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrS 475
 Db 1383 TTGCGTCTACTCTCTCTGTTGTTGCGCAGCTGACCTGGGTATCATCGAATGATACT 1442
 QY 475 erSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspA 495
 Db 1443 CTTCCTCTCTGCTGTCTGCTAGTCGGATGCTGACGGAAGCAATACATCATTTGACG 1502
 QY 495 spPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaA 515
 Db 1503 ATTATCGTAGCTTACTTCTGCTGCTCAGATATCTCTCAGAACGCCAATATCATGT 1562
 QY 515 laTrpTrpAspTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspA 535
 Db 1563 CATGTGGGATTATGGGTATCAATCGTGGCATGGCGACCGCCCAACCTTGGTTGACA 1622
 QY 535 snAsnThrTrpAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluG 555
 Db 1623 ACACACCTGGACACACCATATATGCTACGCTTGGTAGCGATGAGCTCACCGAGG 1682
 QY 555 luLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyG 575
 Db 1683 AAGTCAGCTACCCATCTCCCGCAGCATGATGCTGATTTACGCTGGTGGTGGTTCGGTG 1742
 QY 575 lyLeuIleGlyPheGlyGlyAspIleAsnLysPheLeuTrpMetIleArgIleSerG 595
 Db 1743 GTCTGCTAGGTATCTTGGGATGACATTAAACAAATCTTATGGATGGTCCGATCGCGG 1802
 QY 595 luGlyIleTrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrA 615
 Db 1803 AAGGTATCTGCCCGCATGAGTTAAAGAGCGGAGCTTCTTTACTGCACCGCTGATATC 1862
 QY 615 rgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyr 630
 Db 1863 GTGTGACGATGGAGCGACCCCACTATGCGCAACAGCTTGATAT 1909

RESULT 12

ADB69539

ID ADB69539 standard; DNA; 2882 BP.

XX

AC

XX

XX

DT

XX

DE C. neoformans coding sequence with introns SEQ ID NO:1305.
 XX ds; gene; fungicide; gene therapy; infection.
 KW OS Cryptococcus neoformans.
 XX WO2003052076-A2.
 XX PN 26-JUN-2003.
 XX PD 17-DEC-2002; 2002WO-US040225.
 XX PF 17-DEC-2001; 2001US-0341261P.
 XX PR (ELIT-) ELITRA PHARM INC.
 XX PA Zamudio C, Eroshkin AM;
 XX PI WPI; 2003-533017/50.
 XX DR P-PSDB; ADB70261.
 XX DR PT New nucleic acid, useful for preparing a composition for treating an
 XX PT infection caused by Cryptococcus neoformans.
 XX PS Claim 3; SEQ ID NO 1305; 136pp; English.
 XX CC The invention relates to a novel purified or isolated Cryptococcus
 CC neoformans nucleic acid molecule comprising a sequence encoding a
 CC polypeptide comprising a sequence not given in the specification. A
 CC polynucleotide of the invention has fungicide activity, and may have a
 CC use in gene therapy. The nucleic acid is useful for preparing a
 CC composition for treating an infection caused by Cryptococcus neoformans.
 CC The present sequence represents a C. neoformans sequence of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX SQ Sequence 2882 BP; 615 A; 759 C; 668 G; 840 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,73e-199 Length: 2882
 Score: 2032.50 Matches: 433
 Percent Similarity: 60.86% Conservative: 105
 Best Local Similarity: 48.98% Mismatches: 159
 Query Match: 54.04% Indels: 191
 DB: 10 Gaps: 14

US-10-028-384-6 (1-718) x ADB69539 (1-2882)

QY 13 GlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArg 32
 Db 223 GAAAGCTTGTTCGGCTTCATCATCTCGCGTGAATATATGCGGTGCTCGATTGGAAGTCGA 282
 QY 33 LeuPheAlaValIleLysPheGluSerIleIleHisGluPhe 46
 Db 283 TTGTTTGGGTGATCAGATTGGAATCTGTCATCCACGAATGTGAGTCAAAATCTATCCGC 342
 QY 47 -----AspProTyrPheAsnTyr 52
 Db 343 AGGATTCTCTCAATGCTAACGTCGTGTCTTCATTTCAGTGACCCCGGTTCACATA-GT 401
 QY 53 -----ArgAlaThrLysTyr 57
 Db 402 AAGGCTTCTTCCTGAGGTCAAACTTCCTACTGATTTATTCCTAGCCGAGCTCGAAAGT 461
 QY 57 IleuValAsnAsnSerPheTyrLysPheLeuAsnTyrPheAspArgThrTrpTyrPr 77
 Db 462 TCTTGTAAACAGGGTTTCTACGATCTCGAATCTGGTTCGACCCCTCCCTGGTACCC 521
 QY 77 oLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheI 97
 Db 522 TCTCGGCAGAACTGTCGGTACCAAGCTCTATCTCTGCTGATGTCACGCTCTGACTGAT 581

341 QY -----His-IleProlleleAlas 347
 2274 Db GTCAATGCCATACCTTCGGGATCTATATTATTGTCAACATAGTGCCCATTTATTCCT 2333
 347 QY erValSerGluHisGlnProValSerTTPProAlaphePheAspThrHisPheLeuI 367
 2334 Db CGGTCTCCGACACACACCCCGCTTGGCCCTCATCTACTTTGACCTCGAATGCTTA 2393
 367 QY letRLeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLeuValP 387
 2394 Db TCTCTTTTCCCTGCGGTGCTCTCTGCTGTTTCAAGGAGCTTCGGATGAGCAGATCT 2453
 387 QY heValIleAlaIySerValLeuCysserTyrPheAlaGlyValMetValArgLeuMetL 407
 2454 Db TCATCATCTATTATGCGGTTCCTCAGTGCCATTTTTCGGGTGTCATGGTTCGACTTATGC 2513
 407 QY eutRLeuThrProValIleCyssValSerAlaAlaValAlaLeuSerIyIlePheAspI 427
 2514 Db TTGTCTACAGCGCTGTGTCTGTTTCTCGCCATTGCGTTCCTCCAAACTTCTCGAGG 2573
 427 QY letYrLeuAsp----- 430
 2574 Db CGTATATTGACCCCGTCATCCCGAAAGCGACGAGAGCTGGCGAGTCTCAGACGCGAGG 2633
 431 QY -----PheLystrSerAspAlyLystrAlaIleLysProAla----- 443
 2634 Db TTGTCTCCAGTCCCAAGCGGAGAGAGATGCGCGCTGCCAAGCCCAATAAGAGCGGTTCT 2693
 444 QY -----AlaLeuLeuAlaLysLeuIleValSerGlySerPhe----- 455
 2694 Db CTTTCACAGGTATTTTGGCGGCAAGTCTGTCTCGGCGATCTTTGGTCTCGACACTCGAT 2753
 456 QY -----IlePheTyrLeuTyrLeuPheValPheHisSerThrTrpV 469
 2754 Db TTCTGTGTGTTCATCTCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2813
 469 QY alThrArgThrAlaTyrSerProSerValValLeuProSerGlnThrProAspGlyL 489
 2814 Db TGACTTCAACAGCGTATCTTCGCGCTTCAGTGGTACTTGCATCGCAACCCCGATGGTA 2873
 489 QY ysLeuAlaLeuIleAspAspPheArgGluAlaTyrTyrLeuArgMetAsnSerAspG 509
 2874 Db GCCAAATATCATGTATTTCCGAGAGGCTTACTACTGATTCGCCAAACACCCCGC 2933
 509 QY luAspSerIysValAlaAlaTyrTyrAspTyrGlyTyrGlnIleGlyMetAlaAspA 529
 2934 Db AAGACAGCGTCATCTCTGTGGGATACGGTACCGATCCAGATCGCTGTATGGCTGATC 2993
 529 QY rgThrThrLeuValAspAsnAsnThrTyrAsnAsnThrHisIleAlaIleValGlyLysA 549
 3054 Db GCCCCACCTTGTGTATACATATCTGGATTAACACCCCATATGCCACAGTTGGTAAGG 3053
 549 QY laMetAlaSerProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrV 569
 3054 Db CCATGGCTTCCACGAAGATGTCGATATCTCTATCTTGAGGAAGCATGATGTCGATTACG 3113
 569 QY alLeuValIlePheGlyGlyLeuIleGlyPheGlyAspAspIleAsnlyPheLeuT 589
 3114 Db TTCTTGTGATCTTGGGGCTTATTTGGGCTACTCTGGTGACGATATCAACAGTTTGT 3173
 589 QY rPwIleArgIleSerGluGlyIleTyrProGluGluIleLysGluArgTyrPheTyrT 609
 3174 Db GSATGTTTAGGATCTCACAGGTGAATGGCCCTGACGAGGTGCGAAGATCAACTACTTTA 3233
 609 QY hrAlaGluGlyGluTyrArgValAspAlaArgAlaSerGlu----- 622
 3234 Db CTCAAGAGGGAGTATGTGTGTCATGACAGGGCGTG-CGTCTATTTTTGTATTATTGCC 3292
 623 QY -----ThrMetArgAsnSerLeuLeuTyrL 631
 3293 Db TGGCGAAATTGTGCAATCTGGGACAGCACCCCTACTATGAAGAACTCTCTCATGTACA 3352

631 QY ysMetSerTyrLys----- 635
 3353 Db AAA-GTCTTACTA-CCGGTAGTCAAAATACATTGTGATAGATGAACATCGTCTGACAT 3411
 636 QY -----AspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnM 654
 3412 Db ATCTCAGCTTCCCGAGCTTTATGGTGGACACCCCGCTCAAGACAGGGTTCGAGGCCAAA 3471
 654 QY etIleThrProLeuAspValProLeuAsp-Tyr-Phe----- 666
 3472 Db TTATCCCTCCPAACAGTGT----ACTCTTGATACTCTTGGTAGTATTAGGTTTGAAGAT 3528
 667 QY -----AspGluValPheThrSerGluAsn 674
 3529 Db AAGTGTATAAGACGCTAATGTGTCACATTCTCTGCAGACGAGCGTTCATCCTCGAAAT 3588
 675 QY TrpMetValArgIleTyrGlnLeuLysAspAlaGlnIleClyArgThrLeuArgasp 694
 3589 Db TGGATGTCAGATCTACAAGGTCAGAGAGATCCCATTTGGAGGAGACACACAGGCC 3648
 695 QY Val-----GlyGluLeuThrArgSerSerThrIystrArgArgSerIle 709
 3649 Db GTTACTGCTGGAACGGGGGTAAAGATTGAAGAGAGTCTTAGTCCAGTGGGGCGTG 3708
 710 QY LysArg 711
 3709 Db AAGCG 3714
 RESULT 14
 ADD94789
 ID ADD94789 standard; DNA; 2417 BP.
 XX
 AC ADD94789;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Drosophila melanogaster STT3 gene sequence.
 XX
 KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cystic; immunosuppressive; antitumor therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide;
 XX
 KW fruit fly; gene; ds; STT3.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO2003054008-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 18-DEC-2002; 2002WO-CA001967.
 XX
 PR 20-DEC-2001; 2001US-00028384.
 XX
 PA (COMP-) COMPATIGENE INC.
 XX
 PI Perreault C, Mcbride X;
 XX
 DR WPI; 2003-559122/52.
 XX
 DR P-PSDB; ADD94790.
 XX
 PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 XX Claim 6; SEQ ID NO 7; 66bp; English.
 XX
 XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -

CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, pancreatic cancer, testicular cancer, breast
 CC cancer, melanomas, sarcomas, cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the *Drosophila* melanogaster *SIT3* gene which
 CC is related to the invention. Note: This sequence does not appear in the
 CC specification but was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4,4e-199 Length: 2417
 Score: 2027.50 Matches: 394
 Percent Similarity: 68.22% Conservative: 117
 Best Local Similarity: 52.60% Mismatches: 187
 Query Match: 53.91% Indels: 51
 DB: 10 Gaps: 7

US-10-028-384-6 (1-718) x ADD94789 (1-2417)

QY 12 PheGlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSer 31
 DB 113 TACACAGCCCTATCCATCCGATCTCTGCTAATCGCTGGCTGGCGGATTTCCCT 172
 QY 32 ArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspProTrpPheAsn 51
 DB 173 CGCCTCTTCCGCGTCATCCGTTTCGAGTCGATTCATCCATGAGTTTGCATCGGTTCAC 232
 QY 52 TyrArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnTrpPheAsp 71
 DB 233 TACCGGGCCACCGCTCATGTCGACGATGCTGGTGGTACAACTTCCTCACTGGTTCGAC 292
 QY 72 AspArgThrTrpTyrProLeuGlyArgValThrGlyLysThrLeuTyrProGlyLeuMet 91
 DB 293 GAGCGCGATGATATCCCTCGGAGGATTTGGCGGCTACCGCTCATCCGCGCTGATG 352
 QY 92 ThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIle 111
 DB 353 ATTACGTCGGCGGAATC---CATTGGCTGTCGACGCTACTCAACATACCGGTCCATATT 409
 QY 112 ArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyr 131
 DB 410 CBTGACATCTGGTGTTCCTCGCGCCGATCTTCAGTGGCTGACCTCCATCCATCCACCTAC 469
 QY 132 GluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAla 151
 DB 470 CTGCTGACCAAGAGCTGTGTGCGCGCGCGCTCTTCGCGCGCACTTCATCGCC 529
 QY 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171
 DB 530 ATCGTGGCTGGCTATCAGTAGGTGCGTGGCTGGATCGTACCATACAGGCGCATGCC 589
 QY 172 IleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIle 191
 DB 590 ATATTGGCCCTGCAGTTCACCTACTCTCTGTGGGTGGCTCAGTGAAGCTGATCGCGT 649
 QY 192 MetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyr 211
 DB 650 TTCTGGTGGCGCGAGCGCTTTGTCTTCTTACATGTTGTCCTGGCGGTGGGTGCTAC 709
 QY 212 ValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyr 231
 DB 710 GTGTTTCATCATCACTGATACCCCTGCACGCTCTCTGCTATCTGCTATTATGGCGAGTAC 769

QY 232 SerSerLysLeuTyrSerAlaTyrThrThrTrpTyrAlaIleGlyThrValAlaSerMet 251
 DB 770 TCGCGCGCTGCTCTACCCAGCTACAGACCTTCTACATCTCTGAGCTGTCTCTCATG 829
 QY 252 GlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGly 271
 DB 830 CAGATCCCTTCTGTTGGATTTCACCGATACGACAGTGAACACATGGCTGGCTGGGA 889
 QY 272 ValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThr 291
 DB 890 GTGTTTGTCTCTTATGGCGCTGGCCACTTGGCCGCTTGGAGTTCGAGTCTGGCGCG 949
 QY 292 AlaLysPheLeuValIleMetMetValSerLeuPheLeuIleLeuValValVal 311
 DB 950 AAGGATTCGGAAGCTGTTTCATCGTGGCGGATTCGTTGGTGGCGCTTGGCGTTTGG 1009
 QY 312 GlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrSer 331
 DB 1010 GCGTCTGCTGCTCACCATCTGGCGGTGTTGGCCCGCTGGAGTGCAGCTTCTACTCG 1069
 QY 332 LeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHis 351
 DB 1070 CTGTGGATCTGGCTACGCGCAAGATCCACATTCCTCATCTGATCCGCTGGAGCAT 1129
 QY 352 GlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPhePro 371
 DB 1130 CAGCCACACCTGTTCTCTCTTCTTGTATCTGCACATCTGTTGGTGGCTTCTCCCA 1189
 QY 372 AlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyr 391
 DB 1190 GTGGAGTGTGCTACTGCATCAACAGATCAACGACGAGCGCTTTCGTGGTGTGTAC 1249
 QY 392 SerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrPro 411
 DB 1250 GCATCAGTGGCGTTTACTTCGCTGGTGTGATGTGGCTTGTGATGTGACCTTCACGCG 1309
 QY 412 ValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeuAspPhe 431
 DB 1310 GTGGTGTGATCTGCTGGCGGATGCGCTTTCGGGACTGTGATGTGTTCTCTGCAAGAG 1369
 QY 432 LysThrSerAspArg-----LysTyrAlaIleLys 441
 DB 1370 GATTCGCTTACAGGAATGGGCACAGCCATAGCCACGCCACCGAGTGGATGAAGCTGAG 1429
 QY 437 -----AlaLysLeuIleValSerGlySer 454
 DB 1430 GATTCATTGACAGAGAGAGCTGTACGACAAAGCTGGCAAGCTGAAGCATCTACTAAG 1489
 QY 442 ProAlaAlaLeu-----AlaLysLeuIleValSerGlySer 454
 DB 1490 CATGATGCCACAGCATCTGGGTGCTGCTGCTCAACCTCAAGATGATTTATTTTGGCC 1549
 QY 455 PheIlePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyr 474
 DB 1550 GTTCTAATCTCTGATGATGTTCGCTGCCACTGCGCTGGGTGACAGCAATGCTCTAC 1609
 QY 475 SerSerProSerValValLeuPro---SerGlnThrProAspGlyLysLeuAlaLeuIle 493
 DB 1610 TCCAGTCCCTTCCATGTTCTTGGCTTCCACACAGTCAAGATGGATCCGCCACATTTTA 1669
 QY 494 AspAspPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysVal 513
 DB 1670 GACGATTTTCAGAGAGGCTTACTACTGCTTTCGAGAGAACACTGCGCATGATGCTCGCT 1729
 QY 514 AlaAlaTrpTrpAspTyrGlyTyrGlnIleGlyMetAlaAspArgThrThrLeuVal 533
 DB 1730 ATGCTCTGTGGATTCAGGATACGATACGATACCGGAATGGCAACAGACAGCGTAGTG 1789
 QY 534 AspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerPro 553
 DB 1790 GATTAATAATACGTGAACAATAGTACATAGCTGTTGGCAAGGCAATGTCTTCAACC 1849
 QY 554 GluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePhe 573

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:57:05 ; Search time 139.657 Seconds
(without alignments)
3654.277 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 3761

Sequence: 1 MGSDRSCLVSVFQTLKLV.....TRSTKTRRSIKRPELGLRV 718

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Ygapext 6.0			7.0
Delop 6.0			7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODE=frame+ p2n model -DEV=xlh
-Q=/cgn2_1/USFTO_spool/US10028384/runat 14122004 131520 8052/app_query.fasta_1.3740
-DB=Issued Patents NA -OFMR=fascap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CEN 1.1.259 @runat 14122004 131520 8052 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXI=7

Database : Issued Patents NA:*

1:	/cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2:	/cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3:	/cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4:	/cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5:	/cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6:	/cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3751	99.7	2157	4	US-09-614-221A-318
2	954.5	25.4	867	4	US-09-248-796A-3089
3	892	23.7	1386	4	US-09-270-767-11648
4	891.5	23.7	900	4	US-09-270-767-12331
5	711	18.9	594	4	US-09-248-796A-3090
6	648	17.2	1660	4	US-09-270-767-158
7	648	17.2	1660	4	US-09-270-767-15440
8	595	15.8	507	4	US-09-513-999C-1965
9	591.5	15.7	914	4	US-09-270-767-12856
C 10	509	13.5	1660	4	US-09-270-767-158
C 11	509	13.5	1660	4	US-09-270-767-15440
12	461.5	12.3	503	4	US-09-270-767-27262

13	450	12.0	616	3	US-09-328-111-332	Sequence 332, App
14	409	10.9	487	3	US-09-385-982-213	Sequence 213, App
15	396	10.5	487	4	US-09-702-705-1655	Sequence 1655, App
16	396	10.5	487	4	US-09-736-457-1655	Sequence 1655, App
17	396	10.5	487	4	US-09-614-124B-1655	Sequence 1655, App
18	396	10.5	487	4	US-09-671-325-1655	Sequence 1655, App
19	396	10.5	487	4	US-09-658-824-1655	Sequence 1655, App
20	389	10.3	407	4	US-09-702-705-375	Sequence 375, App
21	389	10.3	407	4	US-09-702-705-1271	Sequence 1271, App
22	389	10.3	407	4	US-09-736-457-375	Sequence 375, App
23	389	10.3	407	4	US-09-736-457-1271	Sequence 1271, App
24	389	10.3	407	4	US-09-614-124B-375	Sequence 375, App
25	389	10.3	407	4	US-09-614-124B-1271	Sequence 1271, App
26	389	10.3	407	4	US-09-671-325-375	Sequence 375, App
27	389	10.3	407	4	US-09-671-325-1271	Sequence 1271, App
28	389	10.3	407	4	US-09-589-184-375	Sequence 375, App
29	389	10.3	407	4	US-09-589-184-375	Sequence 375, App
30	389	10.3	407	4	US-09-658-824-1271	Sequence 1271, App
31	383	10.2	307	4	US-09-702-705-588	Sequence 588, App
32	383	10.2	307	4	US-09-736-457-588	Sequence 588, App
33	383	10.2	307	4	US-09-614-124B-588	Sequence 588, App
34	383	10.2	307	4	US-09-671-325-588	Sequence 588, App
35	383	10.2	307	4	US-09-589-184-588	Sequence 588, App
36	383	10.2	307	4	US-09-589-184-588	Sequence 588, App
C 37	324.5	8.6	560	4	US-03-270-767-1318	Sequence 1318, App
C 38	324.5	8.6	560	4	US-03-270-767-16600	Sequence 16600, App
39	324	8.6	292	4	US-09-313-294A-4442	Sequence 4442, App
40	317	8.4	414	4	US-09-513-999C-271	Sequence 271, App
41	278.5	7.4	269	4	US-09-313-294A-906	Sequence 906, App
42	269.5	7.2	265	4	US-09-313-294A-910	Sequence 910, App
43	268	7.1	291	4	US-09-313-294A-4834	Sequence 4834, App
44	260	6.9	250	4	US-09-513-999C-1438	Sequence 1438, App
45	257	6.8	302	4	US-09-313-294A-6869	Sequence 6869, App

ALIGNMENTS

RESULT 1

US-09-614-221A-318
; Sequence 318, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunanandaa, Balasulojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 318
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-318

Alignment Scores:
Pred. No.: 0
Score: 3751.00
Percent Similarity: 99.86%
Best Local Similarity: 99.86%
Query Match: 99.73%
DB: 4
Length: 2157
Matches: 717
Conservative: 0
Mismatches: 1
Indels: 0
Gaps: 0

US-10-028-384-6 (1-718) x US-09-614-221A-318 (1-2157)

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleuLysLeuValIle 20

Db 1 ATGGGATCCGACCGGTCGTGTGTTTGTCTGTGTTTCAGACCATCTCTAAGCTCGTCATC 60

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3089
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (25)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-248-796A-3089

Alignment Scores:
Pred. No.: 4.16e-106 Length: 867
Score: 954.50 Matches: 178
Percent Similarity: 77.82% Conservative: 36
Best Local Similarity: 64.73% Mismatches: 58
Query Match: 25.38% Indels: 3
DB: 4 Gaps: 1

US-10-028-384-6 (1-718) x US-09-248-796A-3089 (1-867)

QY 446 LeuAlaLysLeuAlaValSerGlySerPheIlePheTyrLeuTyrLeuPheValPheHis 465
DB 37 TTGCAAGAGTTGGTTTACTGACATTTACATTTTACCTTTTCTACITTTGTTTACAT 96
QY 466 SerThrTrpValThrAThrAlaTyrSerSerProSerValValLeuProSerGlnThr 485
DB 97 TGTACTTGGTAAATGCAATGCTTATTCATCCATCAGTGTGTTTATGATCCAGAAAC 156
QY 486 ProAspGlyLysLeuAlaLeuAlaLeuAlaAspPheGluAlaTyrTrpLeuAlaMet 505
DB 157 CCAGATGCTCAACATATCATGATGATTATAGAGACCTTATCTAGTTTAAAGATG 216
QY 506 AsnSerAspGluAlaSerLysValAlaAlaAlaTrpAspTyrGlyTyrGlnLeuGlyGly 525
DB 217 AATACACCAAGATGCTTCCCTGAAGATGTCGTGATGATGATGATGATGATGATGAT 276
QY 526 MetAlaAspArgThrLeuValAspAsnThrTrpAsnThrHisLeuAlaLeu 545
DB 277 ATGGCTGATAGAACCACTTGTGTAACATATCATGGAATACACATATGTCACAT 336
QY 546 ValGlyLysAlaMetAlaSerProGluLysSerTyrGluLeuLeuLysGluHisAsp 565
DB 337 GTTGTAAGCAATGCTTCCCTGAAGATGTCGTGATGATGATGATGATGATGATGAT 396
QY 566 ValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGlyGlyAspAlaLeu 585
DB 397 GTTGATTAATGTTAGTTATTTTGGAGGGTATGAGGTTATGAGGTTATGAGGTTATGAG 456
QY 586 LysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluIleLysGluArg 605
DB 457 AAATCTTATGATGGTAAAGATGCTGAGGATCTGCTGCTGATGATGATGATGATGATGAT 516
QY 606 TyrPheTyrThrAlaGluGlyGlyTyrArgValAspAlaAlaSerGluThrMetArg 625
DB 517 GACTACTTACTGACCGAGGAGATATAAAGTGGATTAAGATGATGATGATGATGATGATGAT 576
QY 626 AsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeuPheAsnGlyGlyGln 645
DB 577 AATCTTTGATGATTAAGTATCTGATCATAGATTCATGATGATGATGATGATGATGATGAT 636
QY 646 AlaThrAspArgValArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyr 665
DB 637 GGTGTTGATAGATTAGAAACCAAAATCCAGCCCAATGAAGTACCGAAATGGAATGTT 696

QY 666 PheAspGluValPheThrSerGluAsnTrpMetValArgIleTyrGlnLeuLysLysAsp 685
DB 697 GTTGAGAGAGCTTTCATCAGAAATTTGATTTGAGAAATTTACAAATTTAAAGATTG 756
QY 686 AspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArgSerSer 702
DB 757 GATTAATGTTGTAGAGATTTACATCAAGCTACTGCTTTTGAAGAATCAATCCGGCACT 816
QY 703 ThrLysThrArgSerIleLysArgProGluLeuGlyLeuArg 717
DB 817 TCCAAAGAAACAGATCCATAAAGAGACCTAAATTCGAAGTAAGA 861

RESULT 3

US-09-270-767-11648
; Sequence 11648, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 11648

; LENGTH: 1386

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-11648

Alignment Scores:
Pred. No.: 3.86e-98 Length: 1386
Score: 892.00 Matches: 172
Percent Similarity: 70.79% Conservative: 51
Best Local Similarity: 54.60% Mismatches: 82
Query Match: 23.72% Indels: 10
DB: 4 Gaps: 4

US-10-028-384-6 (1-718) x US-09-270-767-11648 (1-1386)

QY 15 IleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPhe 34
DB 468 CTGTCAAGCTGGCCATTTCTCATCTCGGACGCGTTTATCATTTGCCACAGCTTGTTC 527
QY 35 AlaValIleLysPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAla 54
DB 528 TCTGTGCTGGATTCGAAAGCGTAATCCATGAGTTCGATCCGTTCACTTCACTACCGCAC 587
QY 55 ThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnTrpPheAspArgThr 74
DB 588 ACGGGTTTCTGGCGGACGAGGCGCTTTTACAGTTCCCACTGCTTCGATGACCGCGCC 647
QY 75 TrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSer 94
DB 648 TGGTATCCCTTGGGCGCCATCATCGGCGGACCATCTATCCCGGCGTGTGTCACCTCG 707
QY 95 AlaPheIleTrpHisAlaLeuArgAsnTrp---LeuGlyLeuProIleAspIleArgAsn 113
DB 708 CGCGCCCTGTACCGCTGATG---TGGCTGCTCAATGTGACCATCGACATACGGAAC 761
QY 114 ValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGluPhe 133
DB 762 GTGTGCTCTCTCGGCGCCCTTCTCTCTCGCTGACCGCTGCTGCTGCTGCTGCTGCTGCT 821
QY 134 ThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleVal 153
DB 822 ACAAAGGAGATACACAGCATGAGGCTGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 881
QY 154 ProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIleThr 173
DB 882 CCCGGTATATCTCTCGATCCGTGGGGGATCGTACGACAAATGAAGGATCGCCATTTTC 941

QY 532 LeuValAspIleThrTrpAsnAsnThrHisIleAlaIleValGlyIysAlaMetala 551
DB 917 CTAGTGGATAATATAGTGGAAACAATAGTACATAGCGCTGGTTGGCAAGCAATGTCT 976
QY 552 SerProGluGluLysSerTrpGluIleLeuLysGluHisAspValAspTrpValLeuVal 571
DB 977 TCACCCGAGGAGAGCTTACGAAATATGACATCTCTTGACGTGGACTAGCTTTGGTG 1036
QY 572 IlePheGlyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheIleTrpMetile 591
DB 1037 ATCTTTCCGCGTGTGATCGCTTATCTCGCATGATATCAACAAGTTCTCTGTGATGGTC 1096
QY 592 ArgIleSerGluGlyIleTrpProGluGluIleLysGluArgTrpPheTrpAlaGlu 611
DB 1097 CGAATTCCTGAGGAGAGCATCCCAAGGACATTTAGGAAGCGATTAATCTTACCGACCGC 1156
QY 612 GlyGluTrpArgValAspAlaArgAlaSerGluThrMetArgAsnSerIleLeuTyIys 631
DB 1157 GGTGAATTCAGGTAGATGCGGAAGTGTCCGCGCTGCTCAACTGCTTATGTACAAA 1216
QY 632 MetSerTrpLysAspPheProGluLeu-----PheAsnGlyGlyGluAlaThrAsp 648
DB 1217 TTAAGTACTACAGATTCGGGGAATTGAAGTTGGACTACAGAGTCCATCTCGGATAGAT 1276
QY 649 ArgValArgGlnMetIleThrProLeuAspValProLeuAspTrpPheAspGlu 668
DB 1277 CGCACAGTAAACCCGCTCATTTGGGAATAAGGACTTC---GATCTGACCTACCTGGAGGAG 1333
QY 669 ValPheThrSerGluAsnTrpMetValA-gileTrpGlnLeuLysLysAsp----- 685
DB 1334 GCCTACACCAACAGACACTGGCTTGTTCATCTATAGGTGAGAGCCGACATGAGTTC 1393
QY 686 -----AspAlaGlnGlyArgTrpLeuArgAspValGlyGluLeuThrArg 700
DB 1394 AATAGACCATCACTGAGACCAAGGAGAGACAGATTCCTCCAGCAAACTTCATTTCCGAGA 1453
QY 701 SerSerThrLysThrArgArgSerIleLysArg 711
DB 1454 AAGAACTCTAAGCTGCGCAAGGCGCTACATACGA 1486

RESULT 8

US-09-513-999C-1965
; Sequence 1965, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.052.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1865
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 122..505
US-09-513-999C-1965

Alignment Scores:

Pred. No.:	9,95e-63	Length:	507
Score:	595.00	Matches:	106
Percent Similarity:	81.44%	Conservative:	30
Best Local Similarity:	63.47%	Mismatches:	31
Query Match:	15.82%	Indels:	0
DB:	4	Gaps:	0

US-10-028-384-6 (1-718) x US-09-513-999C-1965 (1-507)

QY 474 TyrSerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIle 493
DB 2 TACTCTAGTCCCAAGTGTAGTCTCTGGCTCATACATCATGATGATGATGATATCTTA 61
QY 494 AspAspPheArgGluValaTyTrpLeuArgMetAsnSerAspGluAspSerLysVal 513
DB 62 GATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACGAGTA 121
QY 514 AlaAlaTTPtPAspTyTrpGlyGlnIleGlyMetAlaAspArgThrThrLeuVal 533
DB 122 ATGCTCTGGTGGGATTAATGCTATCAGATAGCTGGAATGGCTAATAGAACTACCTGGTG 181
QY 534 AspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerPro 553
DB 182 GATAATACACCTGGAATACAGCCACATAGCATGCTGGTGGAAAGCTATGCTTTCTTAAT 241
QY 554 GluGluLysSerTrpGluIleLeuLysGluHisAspValAspTrpValLeuValIlePhe 573
DB 242 GAAACAGCAGCGCTATAAAATCATCAGGACTCTAGATGATGATGATGATGATGATTTT 301
QY 574 GlyGlyLeuIleGlyPheGlyGlyAspIleAsnLysPheLeuTrpMetIleArgIle 593
DB 302 GGAGGGGTATTTGGCTATCTCTGGTGATGATATCAACAAATTTCTCTGGATGGTTAGATA 361
QY 594 SerGluGlyIleTrpProGluGluIleLysGluArgTrpPheTrpThrAlaGluGlyGlu 613
DB 362 GCTGAAGGAGAGAACATCCCAAGACATTCGGGAAGTCACTATTTACCCACAGGAGAA 421
QY 614 TyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyIysMetSer 633
DB 422 TTCCCTGTAGCAAGACAGCAGGATCCCTACTTTGTGAATTCGCTTATGATATAAATGTCA 481
QY 634 TyrLysAspPheProGlnLeu 640
DB 482 TACTACAGATTTGGAGAAATG 502

RESULT 9

US-09-270-767-12856
; Sequence 12856, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12856
; LENGTH: 914
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-12856

Alignment Scores:

Pred. No.:	7,43e-62	Length:	914
Score:	591.50	Matches:	114
Percent Similarity:	72.59%	Conservative:	29
Best Local Similarity:	57.87%	Mismatches:	43
Query Match:	15.73%	Indels:	11
DB:	4	Gaps:	4

US-10-028-384-6 (1-718) x US-09-270-767-12856 (1-914)

QY 494 AspAspPheArgGluAlaTyTrpLeuArgMetAsnSerAspGluAspSerLysVal 513
DB 5 GATGACTTCCGAGGCGCTACTACTGCTGCAGATGAACACTCCGAGGAGCGCTCGCATA 64
QY 514 AlaAlaTTPtPAspTyTrpGlyGlnIleGlyMetAlaAspArgThrThrLeuVal 533

Db 65 ATGTCCTGGGACTACGGCTACCGATACGATACGCGCATGCGCAATCGGACGATATTAGTG 124
Qy 534 AspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerPro 553
Db 125 GATAACAATACCTGGACACACATATATCGCGGTTGGCCAGGCGATGCTCTTCG 184
Qy 554 GluGlySerTrpGluIleLeuLysGluHisAspValAspTyrValLeuValIlePhe 573
Db 185 GAGGAAGAGCTTACGAGATAATGAGGAACCTGGATGGGACTACGTTCTCGTATTTTC 244
Qy 574 GlyGlyLeuIleGlyPheGlyAspAspIleAsnLysPheLeuTrpMetIleArgile 593
Db 245 GAGGGCTCACTGGCTACTCATCGGACGATATCAACAAGTTCCTGTGGATGGTGGCGATT 304
Qy 594 -----SerGluGlyIleTrpProGluGluIleLysGluAspTyrPheTyrThr 609
Db 305 GCGCGGACGAGATCGTGGT-----GCGCACATCCCGGAAAGGACTACTATGCG 355
Qy 610 AlaGluGlyGlyTyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeu 629
Db 356 GCCAACGGAGAGTTCGAGTGACAAAGGAGGCTCACCCACACACTGCTCAATTGTTTGATG 415
Qy 630 TyrLysMetSerTrpLysAspPheProGluLeuPheAsnGlyGly-----GlnAla 646
Db 416 TACAGATGTGCTACTATCGCTTGGCAATGACAGGAGTGGCAAGCCAGGCG 475
Qy 647 ThrAspArgValArgGlnMetIleThrProLeuAspValProProLeuAspTyrPhe 666
Db 476 TAGCATCGAGTTCGTGCCCGCCGAGATCGGCAACAGGACTTT--GAACCTGGATGCTCGT 532
Qy 667 AspGluValPheThrSerGluAsnTrpMetValArgIleTyrGlnLeuLys 683
Db 533 GAGGAGCGGTACACACGAGGACTGGCTGGTGGCATCTACAAAGTTAG 583

RESULT 10

US-09-270-767-158/c
; Sequence 158, Application US/09270767
; Patent No. 6703491

GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

Alignment Scores:

Pred. No.: 2,35e-51 Length: 1660
Score: 509.00 Matches: 101
Percent Similarity: 68.00% Conservative: 35
Best Local Similarity: 50.50% Mismatches: 52
Query Match: 13.53% Indels: 12
DB: 4 Gaps: 3

US-10-028-384-6 (1-718) x US-09-270-767-158 (1-1660)

Qy 523 IleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHis 542
Db 796 GTTGGCGGAATGCAACAGACGAGCGCTAGTGGATAATATACGTGGAAACAATAGTCAC 737
Qy 543 IleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIleLeuLys 562
Db 736 ATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTCTCTACGAAATATGACA 677
Qy 563 GluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGlyAsp 582
Db 676 TCTCTTGACGTGACTACGTTTGGTGAATCTTTGGCGGTGTGATCGCTATTCCGGCGAT 617

Qy 583 AspIleAsnLysPheLeuTrpMetIleArgileSerGluGlyIleTrpProGluGluIle 602
Db 616 GATATCAACAAGTTCCTGTGGATCGGATCCGAATTCGCGAGGAGAGCATCCCAAGGACATT 557
Qy 603 LysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAlaSerGlu 622
Db 556 AAGGAAGCGATTAATTTACCGACCGCGGTGAATTCAGGGTAGATGCCGAAGTCTCGG 497
Qy 623 ThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu----- 640
Db 496 GCCCTGTCTCACTGCTTATGTACAAATTAAGCTACTACAGATTTCGGGGAATGAAGTTG 437
Qy 641 ---PheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAsp 659
Db 436 GACTACAGAGCCCATCTGGATATGATGCACACGTAACCGCTCATTTGGGAATAAGGAC 377
Qy 660 ValProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgile 679
Db 376 TTT---GATCTGACCTACCTCGGAGGAGGCTACACACAGAACACTGGCTTCTTCGCATC 320
Qy 680 TyrGlnLeuLysLysAsp-----AspAlaGlnGlyArgThr 691
Db 319 TATAGTGTGGAAGAGCGCATGAGTTCATAGACCATCACTGAAGACCAAGGAGAGACG 260
Qy 692 LeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLysArg 711
Db 259 ATTCTCCAGCAAACTTCATCTCGAAGAAAGAACTCGAAGCGTCCGAAGGCTACATACGA 200

RESULT 11

US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent No. 6703491

GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Alignment Scores:

Pred. No.: 2,35e-51 Length: 1660
Score: 509.00 Matches: 101
Percent Similarity: 68.00% Conservative: 35
Best Local Similarity: 50.50% Mismatches: 52
Query Match: 13.53% Indels: 12
DB: 4 Gaps: 3

US-10-028-384-6 (1-718) x US-09-270-767-15440 (1-1660)

Qy 523 IleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHis 542
Db 796 GTTGGCGGAATGCAACAGACGAGCGCTAGTGGATAATATACGTGGAAACAATAGTCAC 737
Qy 543 IleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIleLeuLys 562
Db 736 ATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTCTCTACGAAATATGACA 677
Qy 563 GluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGlyAsp 582
Db 676 TCTCTTGACGTGACTACGTTTGGTGAATCTTTGGCGGTGTGATCGCTATTCCGGCGAT 617
Qy 583 AspIleAsnLysPheLeuTrpMetIleArgileSerGluGlyIleTrpProGluGluIle 602
Db 616 GATATCAACAAGTTCCTGTGGATCGGATCCGAATTCGCGAGGAGAGCATCCCAAGGACATT 557

QY 603 LysGluArgTyrPheTyrThraGluGlyGluTyrArgValAspAlaArgAlaSerGlu 622
DB 556 AAGGAAAGCGAATTAATACCGACCGCGGTGAATTCAGGTAGATCCGAAGGTGCTCG 497
QY 623 ThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu----- 640
DB 496 GCCCTGCTCACTGCTTATGACAAATTAAGCTACTACAGATTCCGGGAATTTGAAGTTG 437
QY 641 ---PheAsnGlyGlnAlaThrAspArgValArgGlnValMetIleThrProLeuAsp 659
DB 436 GACTACAGAGGCCCATCTGGATATGATCGCACACGTAAACCGCTGATCGGAATTAAGGAC 377
QY 660 ValProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgile 679
DB 376 TTT---GATCTGACCTACCTGGAGGAGGCTTACACACAGAACACTGGCTTGTTCGATC 320
QY 680 TyrGlnLeuLysLysAsp-----AspAlaGlnGlyArgThr 691
DB 319 TATAGTGTGAGAACCCCATGATTCATAGACCATCTAGTGAAGACCAAGGAGAGACG 260
QY 692 LeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgSerIleLysArg 711
DB 259 ATTCTCCAGCAAACTTCTATCTCGAAGAAAGAACTCGAAGGCTCGCAAGGCTTACATAGA 200

RESULT 12
US-09-270-767-27262
; Sequence 27262, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; EARLIER FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 27262
; LENGTH: 503
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-27262

Alignment Scores:
Pred. No.: 1,78e-46 Length: 503
Score: 461.50 Matches: 93
Percent Similarity: 70.11% Conservative: 29
Best Local Similarity: 53.45% Mismatches: 45
Query Match: 12.27% Indels: 7
DB: 4 Gaps: 2

US-10-028-384-6 (1-718) x US-09-270-767-27262 (1-503)
QY 155 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleThrLeu 174
DB 2 GGGTATATCTTCGATCGTGGCGGATCTGACAAATGAAGGCAATCCCATTTCTGCG 61
QY 175 LeuMetValThrPheMetPheThrIleLysAlaGlnLysThrLysIleMetHisala 194
DB 62 ATGCTCTTCACTACTATTTGTCATCAAGCGGTAAAGACGGGACGATCTTTTGGTCG 121
QY 195 ThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrValPhele 214
DB 122 GCTATGTCGGCATTTGGCCATCTTCTATGTTCTCTCGTGGGGTGGCTATCTCTCTG 181
QY 215 ThrAsnLeuIleProLeuHisValPheLeuLeuLeuMetGlyArgTyrSerSerLys 234
DB 182 ATTAACCTAATCCCGCTGCACGTCTGGCGCTGATGATCACCGACGTTCTCGCACAG 241
QY 235 LeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIlePro 254
DB 242 ATCTACATAGATACAGACGCTATCTGCTGGCACCATCTTCTGATGATGATCTCG 301
QY 255 PheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGly 274

DB 302 TTGTGGGATTCACACCATCCAGAGCTCCGAACACATGCTGGCACTGGGAACCTTTGCG 361
QY 275 LeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLysPhe 294
DB 362 CTGTGCCAGATTACAGCTTTTGTGACTATCTGCGCTCGCGCATTCACAGGATCACTTC 421
QY 295 LysValIleMetMetValSerIlePheLeuLeuValLeuValGlyValGlyLeuSer 314
DB 422 GATCTG-----CTCTTCAAGACGTTGTGTTTCCAGTGT-----TTGACT 460
QY 315 AlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArg 328
DB 461 GTGGTGTTCGTGGGTACCTGCTCAGCGTTTACCGGAAA 502

RESULT 13
US-09-328-111-332
; Sequence 332, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll, III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 332
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (1)---(616)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-332

Alignment Scores:
Pred. No.: 6,39e-45 Length: 616
Score: 450.00 Matches: 90
Percent Similarity: 65.52% Conservative: 24
Best Local Similarity: 51.72% Mismatches: 54
Query Match: 11.96% Indels: 6
DB: 3 Gaps: 3

US-10-028-384-6 (1-718) x US-09-328-111-332 (1-616)
QY 153 ValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIle 172
DB 5 GTACAGCGCTACATATCTCGGTAGCTGGATCTTTGATATGAAGGCATTGCTATT 64
QY 173 ThrLeuLeuMetValThrPheMetPheThrIleLysAlaGlnLysThrGlySerIleMet 192
DB 65 TTTCACCTTCAGTTCACATACTATTTATGGTAAATCTGTAAAACTGGTCAGTTTTT 124
QY 193 HisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrVal 212
DB 125 TGGACAATGTGCTGCTTATCTTATCTATGCTCTCTCTGCTGCTGCTTATGTA 184

Thu Dec 16 16:25:10 2004

Db	254	GTGGACAATAACACATGGGAATAATACCCATATTTCTCGATAGGCGAGCAATGGCGTCC	313
Qy	553	ProGluGluLysserTyrGluileleuLySGluHisAspValAspTyrValleuValile	572
Db	314	ACAGAGGAAAGCCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGTCGTCGTCATT	373
Qy	573	PheGlyGly	575
Db	374	TTTGAGGA	382

Search completed: December 15, 2004, 11:59:00
Job time : 170.657 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 15, 2004, 06:02:06 : Search time 903.972 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-6
Perfect score: 3761
Sequence: 1 MGSRRSCVSLVVFQTLKLV.....TRSGTKRSIKRPELGLRV 718

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+pz2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10028384/runat_14122004_131523_8140/app_query.fasta_1.3740
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rbp -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-WAXLEN=200000000 -USER=US10028384@cgn_1_1_2577@runat_14122004_131523_8140
-NCPUS=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLCK=100
-JONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3761	100.0	2733	15	US-10-028-384-5
2	3751	99.7	2157	9	US-09-801-368-387
3	3751	99.7	2157	18	US-10-793-639-318
4	2738	72.8	2256	15	US-10-032-585-6323
5	2438.5	64.5	2232	15	US-10-128-714-7139
6	2301.5	61.2	2466	16	US-10-320-797-2305
7	2284.5	60.2	2603	15	US-10-128-714-1339
8	2264.5	60.2	3969	15	US-10-128-714-1339
9	2264.5	60.2	4603	15	US-10-128-714-5139
10	2090	55.6	1848	15	US-10-128-714-2139
11	2042	54.3	1969	15	US-10-128-714-1139
12	2032.5	54.0	2882	16	US-10-320-797-1305
13	2032.5	54.0	4738	16	US-10-320-797-305
14	2032.5	53.9	2417	15	US-10-028-384-7
15	2013.5	53.5	2710	15	US-10-028-384-3
16	2012.5	53.5	2481	15	US-10-028-384-1
17	1946	51.7	3046	18	US-10-417-375-95
18	1945	51.7	3093	18	US-10-417-375-92
19	1945	51.7	3094	15	US-10-028-384-9
20	1934	51.4	5404	18	US-10-417-375-99
21	1934	51.4	5827	18	US-10-417-375-97
22	1920	51.1	2472	15	US-10-171-581-112
23	1920	51.1	2472	15	US-10-028-384-11
24	1920	51.1	2472	15	US-10-172-118-742
25	1920	51.1	2472	16	US-10-342-887-742
26	1754.5	46.6	2681	16	US-10-425-114-1408
27	1754	46.6	2839	18	US-10-425-115-150745
28	1739	46.2	2779	17	US-10-437-963-99904
29	1684	44.8	3197	16	US-10-424-593-111541
30	1680.5	44.7	3141	18	US-10-425-115-130787
31	1402	37.3	2244	17	US-10-437-963-48342
32	1156	30.7	1543	10	US-09-974-879-133
33	1156	30.7	1543	10	US-09-305-736-133
34	1156	30.7	1543	11	US-09-818-683-133
35	1156	30.7	1543	11	US-09-818-683-133
36	1156	30.7	1543	16	US-10-621-401-133
37	1154	30.7	1209	15	US-10-106-698-330
38	1154	30.7	1209	16	US-10-264-237-412
39	1000	26.6	1728	16	US-10-424-593-122476
40	885	23.5	1094	17	US-10-437-963-99902
41	796	21.2	1828	10	US-09-945-527-62
42	739	19.6	1114	16	US-10-296-115-629
43	719.5	19.1	2660	16	US-10-264-049-630
44	661	17.5	887	16	US-10-424-593-77697
45	581.5	14.9	904	17	US-10-767-701-9050

ALIGNMENTS

RESULT 1
US-10-028-384-5
; Sequence 5, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCERIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent version 3.1
; SEQ ID NO 5
; LENGTH: 2733
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: D28952
; DATABASE ENTRY DATE: 1999-02-07

RELEVANT RESIDUES: (1)...(2733)

US-10-028-384-5

Alignment Scores:

Pred. No.: 0
 Score: 3761.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 Ds: 15

Length: 2733
 Matches: 718
 Conservative: 0
 Mismatches: 0
 Indels: 0
 Gaps: 0

US-10-028-384-6 (1-718) x US-10-028-384-5 (1-2733)

Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 Db 219 ATGGATCCGACCGGTGCTGCTGTTTGTCTGTGTTTCAGACCATCTCAAGCTCGTCATC 278
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 279 TTCGTGGCGATTTTGGGGGTGCATATCATCACGTTTGTTCAGTCATCAAAATTTGAG 338
 Qy 41 SerIleIleHisGluPheAspProTTPheAsnTyrArgAlaThrLysTyrLeuValAsn 60
 Db 339 TCATATTCATGATTCGACCCCTGGTTCATATATAGGGCTACCAATATCTCGTCAAC 398
 Qy 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspArgThrTyrTrpProLeuGlyArg 80
 Db 399 AATTTCGTTTACAAAGTTTTCGAACTGGTTGACGACCGTACCTGGTACCCCTCGGAAG 458
 Qy 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla 100
 Db 459 GTTACTCGAGGACTTTATATCTCTGTTGATGACGACTAGTGCCTTCATCTGCACGCC 518
 Qy 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 Db 519 CTGGCAACTGGTGGGCTTGGCCATTGATGACATGACAGAAAGTTTGTGCTATTTGGCGCA 578
 Qy 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 Db 579 CTATTTCTGGGTCACCGCTGGCGACTTACGAAATTTACGAAAGAGATTAAAGATGCC 638
 Qy 141 SerIleGlyLeuLeuAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 Db 639 AGCCCTCGGCTTGGCTGCTGTTTATAGCCATTGTCCTCCCGTTATATATCTAGATCA 698
 Qy 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 699 GTGCGGGTCTTACGATATATGAGCCATTGCCATTACACTATTATATGGTCACCTTCA7G 758
 Qy 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 Db 759 TTTTGGATTAAGGCCCAAAAGATGCTCTATCATGACGCAACGTGTGCGAGCTTTATTC 818
 Qy 201 TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuProLeu 220
 Db 819 TACTTCTACATGCTGCTGGCTGGGTGGATACGTTTCATCAACCACTTGTATCCCATC 878
 Qy 221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 Db 879 CATGCTCTTTTGTGATTTTGA7GGGCGAGATATTCGCCAAACTGTATCTCGCTACACC 938
 Qy 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 939 ACTTGTGTACGCTATTGGAACTGTTGTCATCCATGCGAGATCCCATTTTGTGGTTTCTACCT 998
 Qy 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 999 ATCAGGCTACACACACATGGCCGCTTGGGTGTTTTCGGTTTGAATTCAGATATGTCGCC 1058
 Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 Db 1059 TTCGGTGACTTCGTGAGGGCCAAATCAGCACAGCTTAAGTTTAAAGTTCATCATGATGTT 1118

Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 Db 1119 TCTCTGTTTGTGATCTTGGTCTTGGTGGTGGACTTTCTGCTTGACCTATATATGGGG 1178
 Qy 321 LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle 340
 Db 1179 TTGATTCGCCCTTGGACTGGTAGATTATTCGTTATGGGATACCAACTAGCAAGATC 1238
 Qy 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProAlaPhePhe 360
 Db 1239 CACATTCCTATCATTTGCCCTCCGTTTCGAAACATCAACCCGTTTCGTGGCCGCTTCTTC 1298
 Qy 361 PheAspThrHisPheLeuIleTrpPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 Db 1299 TTTGATACCCACTTTTGTATCTGGCTATTCGCCGCGGTGTAATCTTACTATTTCTTCGAC 1358
 Qy 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
 Db 1359 TTGAAAGACGAGCAGCTTTTGTTCATCGCTTACTCCGTTCTGTGTCGTACTTTGCCGCT 1418
 Qy 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
 Db 1419 GTTATGGTTAGATTGATGTTGACTTTTGACACCACTCATCTGTGTCGCGCGCTGCGCA 1478
 Qy 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
 Db 1479 TTGTCGAAAGATTTTGCATCTACTCTGGATTTCAAGACAAAGTACCGCAAAATACGGCATC 1538
 Qy 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
 Db 1539 AAACTGCGGCACCTACTGCGCAATTTGATTTTCCGATCATTCATCTTTTATTTGTAT 1598
 Qy 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerValVal 480
 Db 1599 CTTTTCGCTTCCATTTCTACTTGGTTAACAGAACTGCACTACTCTTCTCTCTCTGTGT 1658
 Qy 481 LeuProSerGlnTrpProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
 Db 1659 TTGCCATCACAAACCCAGATGTTAAATTTGGCTTTGATCGACACTTCAGGGAAGCGTAC 1718
 Qy 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTyrGly 520
 Db 1719 TATTGGTTAAGAATGAATCTCTGATGAGCAGTAAGGTTTCAGCGTGGTGGGATACCGT 1778
 Qy 521 TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
 Db 1779 TACCAAAATTTGGTCATGGCAGACAGAACCACTTTAGTCGATTAACAACACGTTGGAACAAT 1838
 Qy 541 ThrHisIleAlaIleValGlyAlaMetAlaSerProGluGluLysSerTyrGluIle 560
 Db 1839 ACTCATCTGCCCATCTGTTGGTAAAGCCATGGCTTCCCTTGAAGAGAAATCTTACGAAT 1898
 Qy 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
 Db 1899 CTAAAGAGCATGATGTCGA7TATGCTCTGTCATCTTTGGTGGTCTAAATTTGGTTTGT 1958
 Qy 581 GlyAspPheIleAsnLysPheLeuTrpValIleArgIleSerGluGlyIleTrpProGlu 600
 Db 1959 GGTGATGACATCAACAATTTCTTGTGATGATCAGAATTAGCGAGGAAATCTGCCAGAA 2018
 Qy 601 GluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAla 620
 Db 2019 GAGATAAAGAGCGTTATTTCTATACCGCAGAGGAGATACAGAGTAGATGCAAGGGCT 2078
 Qy 621 SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu 640
 Db 2079 TCTGAGACCATGAGAACTCGTACTTTCAAGATGTTCTACAAAGATTTTCCACAAATTA 2138
 Qy 641 PheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal 660
 Db 2139 TTCAATGTTGGCCAAAGCCACTGACAGAGTGGTCAACAAATGATCATCACCATTAGACGTC 2198
 Qy 661 ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyr 680

```
Db 2199 CCACATTAGACTACTTCGAGAAAGTTTACCTTCGANAACCGATGGTGTAGATATAT 2258
Qy 681 GlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
Db 2259 CAATTGAAGAAGGATGATGTCACCAAGGTAGAACTTTGAGGACGTTGGTGAGTTAACCAAG 2318
Qy 701 SerSerThrLysThrArgArgSerLysLysArgProGluLeuLeuGlyLeuArgVal 718
Db 2319 TCTTCTACGAACACAGAAAGGTCCATAAAGACCTCAATTAGGCTTTGAGAGTC 2372

RESULT 2
US-09-801-368-387
; Sequence 387, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Calli, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofia
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 387
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-387

Alignment Scores:
Pred. No.: 0 Length: 2157
Score: 3751.00 Matches: 717
Percent Similarity: 99.86% Conservatives: 0
Best Local Similarity: 99.86% Mismatches: 1
Query Match: 99.73% Indels: 0
DB: 9 Gaps: 0

US-10-028-384-6 (1-718) x US-09-801-368-387 (1-2157)
Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrLeuLysLeuValIle 20
Db 1 ATGGGATCCGACCGGTGCTGTTGTTGCTGTTTTCAGACCAATCCCTCAAGCTCGTCATC 60
Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
Db 61 TTCGTGGCGATTTTGGGGCTGCCATATCATCACGTTTGTTCAGTCATCAAAATTGAG 120
Qy 41 SerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrLysThrLeuValAsn 60
Db 121 TCTATTATCCATGAATTCGACCCCTGGTTCAATTATAGGGCTACCAAAATATCTCGCAAC 180
Qy 61 AsnSerPheThrLysPheLeuAsnTrpPheAspArgThrTrpTrpProLeuGlyArg 80
Db 181 AATTCGTTTACAAAGTTTGAACCTGGTTTCACGACCGTACCTGGTACCCCTCGGAGG 240
Qy 81 ValThrGlyThrLeuTrpProGlyLeuMetThrThrSerAlaPheIleTrpHisAla 100
```

```
Db 241 GTTACTCGAGGAGCTTTATATCTCTGTTGATGACGACTAGTGGTTCATCTGGCAGGCC 300
Qy 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
Db 301 CTGGCAACTGGTTGGGCTTGCCTACATGACATCAGAAAGCTTTGTGTGCTATTTGGCGCA 360
Qy 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrglyPheThrLysGluLysAspAla 140
Db 361 CTATTTCTGGGTCACCGCTGGGCGACTTACGANTTTACGAAAGAGATTAAAGATGCC 420
Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
Db 421 AGCGTGGGCTTTTGGCTGTGTTTATAGCCATTGTCCTCCGCTTATATATCTAGATCA 480
Qy 161 ValAlaGlySerTyrglyAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
Db 481 GTGGGGGCTCTACGATAATAGGCGCAATGCCATTACACTATTAATGTCACCTTTTCATG 540
Qy 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaLeuPhe 200
Db 541 TTTTGGATTAAAGGCCCAAGAGCTGGCTCTATCATGACGCAACGTGTGCAGCTTTTATTC 600
Qy 201 TyrPheTyrglyMetValSerAlaTrpGlyTyrglyValPheIleThrAsnLeuIleProLeu 220
Db 601 TACTTCTACATGGTTCGGCTTGGGTCGATAGCTGTTTCATCACCAACTTTGATCCCACTC 660
Qy 221 HisValPheLeuLeuIleLeuMetGlyArgTyrglySerSerLysLysLeuTyrglyThr 240
Db 661 CATGCTCTTTTGGCTGATTTTGTATGGGCGAGATATTCGTCCAAAGCTGATTTCTGCACAC 720
Qy 241 ThrTrpTyrglyAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
Db 721 ACTTGGTACGCTATGGAATGTTGATCCATCCATGCGATCCCATTTGCGGTTTCTACCT 780
Qy 261 IleArgSerAsnAspHisMetAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
Db 781 ATCAGGCTTAACGACCACTATGGCGCATTTGGGCTGTTTTCGGTTTGTATCAGATTGTGCGC 840
Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
Db 841 TTCGGTGACTTCGTGAAGGCCCAATCAGCAGCACTAAGTTTAAAGTCATCATGATGGTT 900
Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrgly 320
Db 901 TCTCTGTTTGTATCTTGGTCTTGGTGGTGGTCTTCTGCTGTGACCTATATATGGGG 960
Qy 321 LeuIleAlaProTrpThrGlyArgPheTyrglySerLeuTrpAspThrAsnTyrglyVal 340
Db 961 TTGATTGGCCCTTGGACTGGTAGATTTTATTCGTTATGGATACCACTACCGCAAGATC 1020
Qy 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProAlaPhePhe 360
Db 1021 CACATTCTATCATTTGCTCGCTTTCGAAACATCAACCCGTTTTCGTGGCCCGCTTCTCTC 1080
Qy 361 PheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
Db 1081 TTTGATACCCACTTTTGTACTGCTATTTCCCGCGGGTGATTTCTACTATTTCTCTGAC 1140
Qy 381 LeuLysAspGluHisValPheValIleAlaTyrglySerValLeuCysSerTyrglyPheAlaGly 400
Db 1141 TTGAAGACGAGCAGCAGCTTTTGTCTACTGCTTCTGTTGTCGTACTTTTGGCGGT 1200
Qy 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
Db 1201 GTTATGGTTAGTTGATTTGACTTTTGACACCAAGTTCATCTGTGTGTCGCGCGCGCTCGCA 1260
Qy 421 LeuSerLysIlePheAspIleTyrglyLeuAspPheLysThrSerAspArgLysTyrglyAla 440
Db 1261 TTGTCGAAGATTTTGTACTCTACCTGGAATTCGAAAGACAGTACCGCAAAATACGCAATC 1320
Qy 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrglyLeuTyrgly 460
Db 1321 AAACCTGGGCACTACTGGCCAAATGATTTGTTTCCGATCATCTATTTATTTGAT 1380
```

QY 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaThrSerSerProSerValVal 480
 DB 1381 CTTTTCGCTTCCATCTTACTTGGTAACAAGAACTGCATACCTCTCTCTCTCTCTCTT 1440
 QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuLeuLeuLeuLeuLeuLeuLeu 500
 DB 1441 TTGCCATCAACAAACCCAGATGTTAAATTGGCGTTCATCGACGACTTCAGGAGCGTAC 1500
 QY 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaAlaAlaAlaAlaAla 520
 DB 1501 TATTGGTAAAGAACTCTGATGAGGACGAGTAAGTTGCAGCGTGGTGGGATTAACGCT 1560
 QY 521 TyrGlnLeuGlyGlyMetAlaAspArgThrThrLeuValAlaAspAsnAsnThrTrpAsnAsn 540
 DB 1561 TACCAAAATGGTGCGATGGCAGACAGAACCACTTTAGTGCATAACAACAGCGTGAACAAT 1620
 QY 541 ThrHisLeuAlaLeuValGlyLysAlaMetAlaSerProGluGluLysSerTyrcLuuile 560
 DB 1621 ACTCACATCGCCATCGTGTGTAAGCCATGGCTTCCCTGGAAGAGAAATCTTACGAAAT 1680
 QY 561 LeuLysGluHisAspValAspTyrValLeuValilePheGlyGlyLeuileGlyPheGly 580
 DB 1681 CTAAGAGAGCATGATGCGATTATGCTTGGTTCATCTTGGTGGTCTAATTGGGTTGGT 1740
 QY 581 GlvAspAspIleAsnLysPheLeuTrpMetIleAspGlySerGluGlyLysLeuTrpProGlu 600
 DB 1741 GGTGATGACATCAACAAATCTTGTGGATGATCAGAAATAGCGAGGGAATCTGGCCAGAA 1800
 QY 601 GluileLysGluArgTyrPheThrAlaGluGlyGlyLysArgValAlaAspAlaArgAla 620
 DB 1801 GAGATAAAGAGCGTGATTTCTATACCGCAGAGGAGAGATACAGAGTAGATCAAGGGCT 1860
 QY 621 SerGluThrMetArgAsnSerLeuLysMetSerLysLysAspPheProGlnLeu 640
 DB 1861 TCTGAGACCATGAGAACTCGTACTTTACAGATGCTCCCAAGATTTCCCAAAATTA 1920
 QY 641 PheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal 660
 DB 1921 TTCAATGGTGCCCAAGCCACTGACAGAGTGGTCAACAAATGATCACACCATTAACGCTC 1980
 QY 661 ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyr 680
 DB 1981 CCACATAGACTACTTCGACGAAGTTTTTACTTCGAAAACCTGATGATGATGATGATGAT 2040
 QY 681 GlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
 DB 2041 CAATTGAAGAAGGATGATGCCAAGGTAGAACTTTGAGGAGCGTGGTGGTGAACACAGG 2100
 QY 701 SerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLysArgVal 718
 DB 2101 TCTTCTAGAAACCAAGAGTCCATAAAGAGACCTGAATAGGCTTGAGAGTC 2154

RESULT 3

US-10-793-639-318
 ; Sequence 318, Application US/10793639
 ; Publication No. US20040199940A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Karunandaa, Balasulojini
 ; APPLICANT: Yu, Jaehyuk
 ; APPLICANT: Kishore, Ganesh M.
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
 ; WITH STEROL SYNTHESIS AND METABOLISM
 ; FILE REFERENCE: 16516.075
 ; CURRENT APPLICATION NUMBER: US/10793,639
 ; PRIORITY FILING DATE: 2004-03-05
 ; PRIORITY FILING DATE: US/09/614,221A
 ; PRIORITY FILING DATE: 2000-07-11
 ; PRIORITY FILING DATE: 1999-07-12
 ; NUMBER OF SEQ ID NOS: 626
 ; SEQ ID NO 318
 ; LENGTH: 2157

TYPE: DNA
 ORGANISM: Saccharomyces cerevisiae
 US-10-793-639-318

Alignment Scores:

Pred. No.: 0 Length: 2157
 Score: 3751.00 Matches: 717
 Percent Similarity: 99.86% Conservative: 0
 Best Local Similarity: 99.86% Mismatches: 1
 Query Match: 93.73% Indels: 0
 DB: 18 Gaps: 0

US-10-028-384-6 (1-718) x US-10-793-639-318 (1-2157)

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValile 20
 DB 1 ATGGGATCCGACCGGCTGGTGTGTTTGTCTGTGTTTCAGACCACTCTCAAGCTCGTCATC 60
 QY 21 PheValAlaIlePheGlyValAlaAlaIleSerSerArgLeuPheAlaValileLysPheGlu 40
 DB 61 TTCGTGGCGATTTCCTGGGCTGCCATATCATCAGCTTGTTCGAGTCATCAAAATTTGAG 120
 QY 41 SerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrLysTyrLeuValAsn 60
 DB 121 TCTATTATCATGAATTCGACCCCTGGTTCAATTATAGGCTACCAAAATATCTCGTCAAC 180
 QY 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspAspArgThrTrpTyrProLeuGlyArg 80
 DB 181 AATTGGTTTACAAGTTTTTGAACCTGGTTGACGACCGTACCTGGTACCCCTCGGAAGG 240
 QY 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrPheAla 100
 DB 241 GTTACTGGAGGACATTTATATCTGTTTGTATGACGACTAGTGGTTCATCTGGCACGCC 300
 QY 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 DB 301 CTGCGCAACTGGTGGGCTTGCCCATTCACATCAGAAACGTTTGTGTGCTATTTCGCCCA 360
 QY 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 DB 361 CTATTTCCTGGGCTCACCCTGGCGACTTACGAAATTTACGAAAGAGATTAAGATGCC 420
 QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 DB 421 AGCCTGGGCTTTTGGCTGCTGTTTATAGCAATGTCCTCCCGTGTATATCTAGATCA 480
 QY 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 DB 481 GTGGCGGGGCTCTACGATAATGAGGCCATTCGCATTTACACTATTAAATGGTCACCTTCATG 540
 QY 181 PheTrpIleLysAlaGlnLysThrClySerIleMethIleAlaThrCysAlaAlaLeuPhe 200
 DB 541 TTTTGATTAAAGCCCAAGAGCTGCTCATATGACGACCACTGTCGAGCTTTATTC 600
 QY 201 TyrPheTyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeu 220
 DB 601 TACTTCTACATCGTCTGGCTTGGGTGGATACGCTTTCATCAACCACTTCATCCCACTC 660
 QY 221 HisValPheLeuLeuIleLeuMetClyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 DB 661 CATTCTTTTCTGATTTTATGGGCGCATATTCGTCGCAACCTGATTCGCTACACC 720
 QY 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 DB 721 ACTTGGTACGCTATTGGAACTGTTGCATCCATGAGATCCCATTTGTCGGTTCCTACCT 780
 QY 261 IleArgSerAspAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 DB 781 ATCAGCTCTAACGACCACTGCGCCCATTTGGTGTGTTTCGTTTGCATCAGATTGTCGCC 840
 QY 281 PheGlyAspPheValLysGlyLysIleSerThrAlaLysPheLysValileMetMetVal 300
 DB 841 TTCGTGATCTTCGTGAAGGGCCAAATCATCAGCAGCTAAGTTAAAGTCATCATCATGATGTT 900

QY 301 SerLeuPheLeuLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
Db 901 TCTCGTTTTCGATCTTGGTCTGCTGGTGTGCGACTTCTGCTTCGACCTATATGGG 960
QY 321 LeuLeuAlaProTirThrGlyArgPheTyrSerLeuTirPheThrAsnTyrAlaValSile 340
Db 961 TTGATTGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCACTACGCAAGATC 1020
QY 341 HisLeuProLleAlaSerValSerGluHisGlnProValSerTirProAlaPhePhe 360
Db 1021 CACATTCTCATCTTGGCTCGCTTTCGAAACATCAACCCGTTTCGTCGCCGCTTCTTC 1080
QY 361 PheAspThrHisPheLeuLeuTirPheProAlaGlyValPheLeuLeuPheLeuAsp 380
Db 1081 TTTGATACCACTTTTGTATCTGGCTATTCCCGCGGTGTATCTCTACTATTCTTCGAC 1140
QY 381 LeuLysAspGluHisValPheValLleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
Db 1141 TTGAAGAGCAGCAGCTTTTGTATCGCTTACTCCGTTCTGTGTTCTGTTACTTTGGCGGT 1200
QY 401 ValMetValArgLeuMetLeuThrLeuThrProValLleCysValSerAlaAlaValAla 420
Db 1201 GTATGGTTAGATTGATGTTGACTTTGACACAGTTCATCTGTGTGTCGCGCGCTGCA 1260
QY 421 LeuSerLysLlePheAspLleTyrLeuAspPheLysThrSerAspArgLysTyrAlaLle 440
Db 1261 TTGTCCAGATATTGACATCTACCTGGATTTCAGACAAAGTACCCCAATACGCCATC 1320
QY 441 LysProAlaAlaLeuAlaLysLeuLleValSerGlySerPheLlePheTyrLeuTyr 460
Db 1321 AAACCTGGCGCACTACTGGCAAAATGATTGTTCCGGATCATCTATTATTGAT 1380
QY 461 LeuPheValPheHisSerThrTirPheValThrArgThrAlaTyrSerProSerValVal 480
Db 1381 CTTTTCGCTTCCATCTACTTGGTAAAGAACTGCTGCTCTCTCTCTCTCTCTCTCTCT 1440
QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuLleAspAspPheArgGluAlaTyr 500
Db 1441 TTGCCATCACAAACCCAGATGGTAAATTTGGCTGTATCGACGACTTCAGGAAGCGTAC 1500
QY 501 TyrTirPheLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTirPheAspTyrGly 520
Db 1501 TATTTGGTTAAGAAATGAATCTGTATGAGACAGTAAAGTTGCGAGCTGGTGGGATACGGT 1560
QY 521 TyrGlnLleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTirPheAsn 540
Db 1561 TACCAAAATTTGGTGGCATGGCAGACAGAACCACTTTAGTCGATAACAACAGTGGAAAT 1620
QY 541 ThrHisLleAlaLleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluLle 560
Db 1621 ACTCATCGCCATCGTTGGTAAAGCCATGGCTTCCCGTGAAGAGAAATCTTACGAAAT 1680
QY 561 LeuLysGluHisAspValAspTyrValLeuValLlePheGlyGlyLeuLleGlyPheGly 580
Db 1681 CTAAAGAGCATGATGTGCTATGCTTGTGTGCTATCTTTGGTGTCTTAATGGTTGGT 1740
QY 581 GlyAspAspLleAsnLysPheLeuTirPheMetLleArgLleSerGluGlyLleTirProGlu 600
Db 1741 GGTGATGACATCAACAAATCTTTGGATGATCAAGATTACGAGGAAATCTTCCGCAAA 1800
QY 601 GluLleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAla 620
Db 1801 GAGATAAAGAGCGGTATTTCTATACCGCAGAGGAGAAATACAGATAGATGCAAGGCT 1860
QY 621 SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu 640
Db 1861 TCTGACCATCAGGAATCTCGTACTTTTACAAGATGTCCTACAAGATTTCCCAACATTA 1920
QY 641 PheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetLleThrProLeuAspVal 660
Db 1921 TTCAATGGTGGCCAGCCACTGACAGAGTGGGTCAACAAATGATCACACCATTAGAGTCT 1980

QY 661 ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgLleTyr 680
Db 1981 CCACCATTAGACTACTTCGACGAAAGTTTATTCGAAACATCGGATGGTTAGAAATATAT 2040
QY 681 GlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
Db 2041 CAATTGAAGAGGATGATGCCAAGGTAGAACCTTGAGGACGCTTGGTGAGTTAACCAAG 2100
QY 701 SerSerThrLysThrArgArgSerLleLysArgProGluLeuGlyLeuArgVal 718
Db 2101 TCTTCTAGAAACAGAGGTCATTAAGAGACCTGAATTAGCTTGAGAGTC 2154
RESULT 4
US-10-032-585-6323
; Sequence 6323, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Ho, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6323
; LENGTH: 2256
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6323
Alignment Scores:
Pred. No.: 2,77e-274 Length: 2256
Score: 2738.00 Matches: 509
Percent Similarity: 81.22% Conservative: 88
Best Local Similarity: 69.25% Mismatches: 110
Query Match: 72.80% Indels: 28
DB: 15 Gaps: 3
US-10-028-384-6 (1-718) x US-10-032-585-6323 (1-2256)

QY 9 LeuSerValPheGlnThrLleLeuLysLeuValLlePheValAlaLlePheGlyAlaAla 28
Db 52 GTTGAACCTGATTAGAGTTTATTAAAGATTATTATATATATATATATATATATATATAT 111
QY 29 IleSerSerArgLeuPheAlaValLleLysPheGluSerLleLleGluPheAspPro 46
Db 112 ATTTCTCTCTTTTATTTTCCGTTGATTCGATTTGAAAGATTATTATTCATGATTCGATCT 171
QY 49 TrpPheAsnTyrArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsn 68
Db 172 TGGTTCAATTTCCGAGCAACCAATATTTAGTCACCTCATCTTTTATGAAATTTTGAAT 231
QY 69 TrpPheAspArgThrTirPirProLeuGlyArgValThrGlyGlyThrLeuTyrPro 88
Db 232 TGGTTTGNATAGAACTTGGTACCCTTGGGAGAGTCACTGCTGGTACTTTATATATCC 291
QY 89 GlyLeuMetThrThrSerAlaPheLleTirPheAlaLeuArgAsnTrpLeuGlyLeuPro 108
Db 292 GGTATAATGGTGGCTTCCAGTGCCATTTGGCATATTTTACGTGATTCGTTTGCCTTACCC 351
QY 109 IleAspLleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTirP 128
Db 352 GTTGATTAAGAAATATTGTGTTTATAGCACCAAGTTTCTCGGATTAACGCAAT 411
QY 129 AlaThrTyrGluPheThrLysGluLleLysAspAlaSerAlaGlyLeuLeuAlaAlaGly 148
Db 412 TGTACTTATTTTGGACTAAGAAATGAGGATTTCTAGTCAGGATTTATTGGCAGCTATA 471
QY 149 PheLleAlaLleValProGlyTyrLleSerArgSerValAlaGlySerTyrAspGlu 168
Db 149 PheLleAlaLleValProGlyTyrLleSerArgSerValAlaGlySerTyrAspGlu 168

Db 472 TTTATGGGATGCCCCAGGTTATATTCAAGATCAGTGGCTGCTTCTTATGATTAATGAA 531
Qy 169 AlalileAlaileThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThr 188
Db 532 GCAATTGTCATTACTTTTATTAAATGCGCAACATTTTATTCTGGATTAAATCAATGAATG 591
Qy 189 GlySerIleMetHisAlaThrCysAlaAlaLeuPheThrPheMetValSerAlaTrp 208
Db 592 GGTTCAGTTTCTATGCAACATTCAGACGCAATTTCTATTCTATAGTTAGTCTGG 651
Qy 209 GlyGlyTrpValPheIleThrAsnLeuLeuPheLeuHisValPheLeuLeuLeuMet 228
Db 652 GGTGGATATGTTTCAATACCAATTCATTCATTCATTCATTCATTCATTCATTCATG 711
Qy 229 GlyArgTrpSerSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuVal 248
Db 712 GGTGTTTAAATGCAACATTTTCACTGCTTATCTATCTATCTATCTATCTATCTATCT 771
Qy 249 AlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAla 268
Db 772 GCATCAATCGAGATTCATTCGTTGGTGTTCATTCATTCATTCATTCATTCATTCATG 831
Qy 269 AlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyValPheValLysGln 288
Db 832 GCATPAGGATTTTGGATGTTTCAATAGTGGCTTTTGGTATTCATTCATTCATTCATTC 891
Qy 289 IleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuLeuValLeu 308
Db 892 GTTCCACCAACCAATTTAAATCATCTCTGATAGTTTCCATTCATTCATTCATTCATTC 951
Qy 309 GlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProThrPheGlyArg 328
Db 952 GGTATGTTGGATTTTGGATTTACAGCAATGGTGGATTCCTCTTGGACAGGTACA 1011
Qy 329 PheTrpSerLeuTrpAspThrAsnTrpAlaLysIleHisIleProIleIleAlaSerVal 348
Db 1012 TTTTATTCCTTATGGATACAAATATGCAAGATTCATTCATTCATTCATTCATTCATTC 1071
Qy 349 SerGluHisGlnProValSerTrpProAlaPhePheAspPheHisPheLeuLeuTrp 368
Db 1072 TCTGACATCACTACTGCTGTCGACCAATTCCTTTTCTGACTAGTATGCTTATGG 1131
Qy 369 LeuPheProAlaGlyValPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuVal 388
Db 1132 TTTATTCCTGCTGCTATCTATTTATGTTTCCAAAGATTCGAAGATTCGAACGCTTTTCAT 1191
Qy 389 IleAlaTrpSerValLeuCysSerTrpPheAlaGlyValMetValArgLeuMetLeuThr 408
Db 1192 ATCATTTACAGTATTTGTTCTTATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1251
Qy 409 LeuThrProValIleCysValSerAlaValAlaLeuSerLysIlePheAspIleTrp 428
Db 1252 TTGACTCCAGTCATTTGTTGCTGTCGACCAATTCCTTATCTAAATGTTTGTATGCTAT 1311
Qy 429 LeuAspPhe----- 431
Db 1312 TTGGACATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1371
Qy 432 -----LysThrSerAspArgLysTrpAlaLysProAlaLeu 445
Db 1372 GAATCCAAAGCTCAACCAAAATCAAGTTCAGATTTCCATTT-----GCTGGATAT 1425
Qy 446 LeuAlaLysLeuIleValSerGlySerPheIlePheTrpLeuTrpLeuPheValPheHis 465
Db 1426 TTGTCAAAAGTTTGGTTTACTGACATTTTACATTTTACCTTTTCTTCTTCTTCTTCTTCT 1485
Qy 466 SerThrTrpValThrArgThrAlaTrpSerSerProSerValValLeuProSerGlnThr 485
Db 1486 TGTACTTGGTAAACATCGAATGCTTATTCATCCATCAGTTGTTTGTAGCATCCAGAAC 1545
Qy 486 ProAspGlyLysLeuAlaLeuIleAspPheArgGluAlaTrpTrpLeuArgMet 505
Db 1546 CCAGATGCTCACACATATCATGATGATATAGAGAGCCCTTATTCGTTAAGATG 1605

Qy 506 AsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTrpGlyTrpGlnIleGlyGly 525
Db 1606 AATACACAGAGAGATGCCAAAGTTATGCGCTGGTGGATTTATGTTTATCAATCGGGGT 1665
Qy 526 MetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIle 545
Db 1666 ATGGCTGATAGAACACACTTGTGATACCAATACATGGAATACACACATATTCCTACT 1725
Qy 546 ValGlyValAlaMetAlaSerProGluLysSerTrpGluIleLeuLysGluHisAsp 565
Db 1726 GTTGGTAAGGCAATGCTTCCCTGAAGATGCTGATGAATTTTGAGACACACAGAT 1785
Qy 566 ValAspTrpValLeuValIlePheGlyLeuIleGlyPheGlyGlyAspAspIleAsn 585
Db 1786 GTTGATTAAGTGTAGTATATTGGAGGTTATGGGTTATTCCTGATGATATTAAC 1845
Qy 586 LysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluIleLysGluArg 605
Db 1846 AATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1905
Qy 606 TyrPheTrpAlaGluGlyLysTrpArgValAlaAspAlaArgAlaSerGluThrMetArg 625
Db 1906 GACTACTTTACTGACCGAGGAGATATAAGTGGATAAGATGATCCTGCGCAATGAAG 1965
Qy 626 AsnSerLeuLeuTrpMetSerTrpLysAspPheProGlnLeuPheAsnGlyGln 645
Db 1966 AATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2025
Qy 646 AlaThrAspArgValArgGlnGlnMetIleThrProLeuAspValProLeuAspTrp 665
Db 2026 GGTGTTGATAGATTTAGAAACCAACATCCAGCCATGAGTACCGAATTCATGTT 2085
Qy 666 PheAspGluValPheThrSerGluAsnTrpMetValArgIleTrpGlnLeuLysLysAsp 685
Db 2086 GTTGAGAAAGCCCTTCACATCAGAAAATTTGATGAGATTTTCAAAAGTTAAAGATTG 2145
Qy 686 AspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArgSerSer----- 702
Db 2146 GATAATGTTGTTGATAGATTTTACATCAAGCTACTGCTTTTGAAGATCATTCGGCACT 2205
Qy 703 ThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArg 717
Db 2206 TCCAAAAGAACAGATCCATTAAGACCTAAATTTGAAGTAAGA 2250

RESULT 5
US-10-128-714-7139
Sequence 7139, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Weng
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroskin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
METHODS OF USE
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: PatentIn version 3.1

```

; SEQ ID NO 7139
; LENGTH: 2232
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-7139

Alignment Scores:
Pred. No.: 4,57e-242 Length: 2232
Score: 2428.50 Matches: 460
Percent Similarity: 77.68% Conservative: 97
Best Local Similarity: 64.16% Mismatches: 137
Query Match: 64.57% Indels: 23
DB: 15 Gaps: 6

US-10-028-384-6 (1-718) x US-10-128-714-7139 (1-2232)

QY 15 IleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPhe 34
DB 61 CTTCTCGGATATTATCTGTGTACCATTCAGCAGCAGAGTGGCGTGGACTCTTC 120
QY 35 AlaValIleLysPheGluSerIleIleHisGluPheAspProTTPPheAsnTyrArgAla 54
DB 121 AGCGTTATCGGTTTCAGAGTATCATCCAGAAAGTTGACCGTGGTTCACACTCCGAGCA 180
QY 55 ThrLysTyrLeuValAsnSerPheTyrLysPheLeuAsnTTPPheAspArgThr 74
DB 181 ACAANAATCTTAGTACAGAAATGTTTCTATAGCTTTTGGGATTTGGTTGATGACGGAACA 240
QY 75 TTPtyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSer 94
DB 241 TGGCATCTCTCGGACGTGTCAACCGGTGGCAGCTTATATCCGGTCTCATGGTGACGAGC 300
QY 95 AlaPheIleThrHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnVal 114
DB 301 GCGTGATCTACCATCTCTTGGCA--TTCTTACTATCCCGCTCGATATTCGCAACATC 357
QY 115 CysValLeuPheAlaProLeuPheSerGlyValThrAlaTTPAlaThrTyrGluPheThr 134
DB 358 TGGCTCTACTTGGCGCAGGATCTCCGCTGACTGCATTTGGCAATGTACTTCTCTGACA 417
QY 135 LysGluIleLysAspAla--SerAlaGlyLeuLeuAlaIleGlyPheIleAlaIleVal 153
DB 418 TCGAGATCTCTTCGCACTCTCGAGTCTTCTGCGAGCTTCTATCGGAGCTTTCATGGGAATGCC 477
QY 154 ProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIleThr 173
DB 478 CTGGGTATCATCTCCGATCAGTCTGCTGAAAGTACGATAACGAAGCGATTTGCCATCTT 537
QY 174 LeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMetHis 193
DB 538 CTGCTTGTTTCACATCTTTCTATGGATCAAGGCTGTCAAAAATGGGCTTATCATGTGG 597
QY 194 AlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTTPGlyGlyTyrValPhe 213
DB 598 GGAGCGCTACCGCACTATTCTACGGCTACATGGTGTGGCATGGGTGGGTATGCTTCTC 657
QY 214 IleThrAsnLeuLeuProLeuHisValPheLeuLeuLeuMetGlyArgTyrSerSer 233
DB 658 ATTACGAACCTGATCCCTCTGACCGTTTGTGCTTCTGTGCTGCGGTGATGATGACGACT 717
QY 234 LysLeuTyrSerAlaTyrThrTTPtyrAlaIleGlyThrValAlaSerMetGlnIle 253
DB 718 CGCATCTACATAGTATATACCATGTTATGGTATGGTGGGACTTTTGGCTAGCATGCAGATT 777
QY 254 ProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGlyValPhe 273
DB 778 CCCTTCTCGGATTTTGGCTATCCGAACACGACGACCATGTCCGCTTGGGGTCTTC 837
QY 274 GlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLys 293
DB 838 GGCCTGCTTCAAGTGTGGCTTCCGCGAGTTTGTCCGAGCTTGTCCGAGCTTGTCCAAAGCAG 897
QY 294 PheLysValIleMetMetValSerLeuPheLeuLeuValLeuGlyValValGlyLeu 313

```

```

DB 898 TTCCAGAGACTCTGACCGCATGATCTTCATCACCTTCGGTCTCGTTTGGGCTA 957
QY 314 SerAlaLeuThrTyrMetGlyLeuAlaProTTPThrGlyArgPheTyrSerLeuTyr 333
DB 958 GTTGTCTTCTGACTGTGACGGGAGTATCGCTCTTGGAGCGCGGATTTCTACTCTTGTGG 1017
QY 334 AspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnPro 353
DB 1018 GACACTGGCTATGCCAAATCCATTCCTCATCATTCCTCAGTCTCGGAACACGAGCC 1077
QY 354 ValSerTTPProAlaPhePheAspThrHisPheLeuIleTTPLeuPheProAlaGly 373
DB 1078 ACCGCTTGGCAGCGTCTTCTTCGATCTGAATCTGATCTGGCTTTTCCCGGAGGT 1137
QY 374 ValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyrSerVal 393
DB 1138 GTTACATGTGCTTCCGTGACCTCAAGGACGAGATGTCTTCGTCAATATCTACTCGGTC 1197
QY 394 LeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValIle 413
DB 1198 CTTGCGAGCTACTTCCGCGGTGTTATGTCGCGACTAATGCTGACCTTGACCCCTATTGTG 1257
QY 414 CysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu----- 429
DB 1258 TGTGTTCCGCTGCTTGGCGCTGTCTGCATCTCTGCACACCTATATGGCGACTACCCCTC 1317
QY 430 -----AspPheLysThrSerAsp----- 435
DB 1318 CGGACACGAGCGTCTGAAGCGGAAACGAATGAAGACTCGCTTTCACAACTCTTCGCTCA 1377
QY 436 ---ArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySer 454
DB 1378 GTTCCGAAGCCCAATGTTGGAATCACCTCCCATGTTTCTAAGATATATAGTACGCGGTCT 1437
QY 455 PheIlePheTyrLeuTyrLeuPheValPheHisSerThrTTPValThrArgThrAlaTyr 474
DB 1438 GTTGTGCTTACCTGCTCTGTTGTTGCGCACTGCACTGGGTTCATCGAATGCATAC 1497
QY 475 SerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAsp 494
DB 1498 TCTTCTCTCTCCGTGCTTGGCTAGTCCGATGCTGACGGAAGCCAAATACATCATGTAC 1557
QY 495 AspPheArgGluAlaTyrTyrTTPLeuArgMetAsnSerAspGluAspSerLysValAla 514
DB 1558 GATTATCGTAGGCTTACTACTGCTTCGTGAGAAATCTCTCAGAACGCGCAAAATCATG 1617
QY 515 AlaTTPTrpAspTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAsp 534
DB 1618 TCATGTGGGATTTATGGGTATCAATCGTGGCATGGCGACCGCCCAACCTTGGTTGAC 1677
QY 535 AsnAsnThrTTPAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGlu 554
DB 1678 AACAACTCGAACACACACCATATTTGCTACGTTGGTAAGCGCATGAGCTCACGCGAG 1737
QY 555 GluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGly 574
DB 1738 GAAGTCAGCTACCCCATCTCCGCGCATGATGTCGATACGCTGCTGGGTGGTTCGGT 1797
QY 575 GlyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuTTPMetIleArgIleSer 594
DB 1798 GGTCTGCTAGGTATTCTGCGCATGACATTAACAAATTTCTATGGATGGTCCGCTATCGCC 1857
QY 595 GluGlyIleTTPProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyr 614
DB 1858 GAAGTATCTCGCCCGATGAGGTAAAGAGCGGAGCTTCTTTACTGCACGCGGTGAATAT 1917
QY 615 ArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyr 634
DB 1918 CGTGTGACGATGGAGCGACCCCACTATCGCAACAGCTTGTGGTATATAAATGCTTAT 1977
QY 635 LysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnMet 554

```


Db 1978 TACAATTTCAACTCTCTCTCCGTCGGCCCAAGCTGTGACCGCTCCGTCGGTCAAAA 2037
 Qy 655 IleThrProLeuAspValProLeuAspTyrPheAspGluValPheThrSerGluAen 674
 Db 2038 CTT--CCACAGAGAGGCCCTCAGCTCTCTACACTCGAAGAAGCTTTTCCAGCGAGAAC 2094
 Qy 675 TtpMetValArgIleTyrGluLeuLysAspAlaGlnGlyArgThrLeuArgAsp 694
 Db 2095 TGGATCATCTGATCTACTCAAGTCAAGTCTTGACCACTTGGCCGAGACCAACAC 2154
 Qy 695 ValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLysArg 711
 Db 2155 GCTGTTGCTTCCGACAAA---GGTCTCAAGAAAAGCGGAGTACAAAGAGG 2202

RESULT 6

US-10-320-797-2305
 ; Sequence 2305, Application US/10320797
 ; Publication No. US20040014955A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eroskin, Alexey M.
 ; APPLICANT: Zamudio, Carlos
 ; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
 ; FILE OF INVENTION: METHODS OF USE
 ; FILE REFERENCE: 10182-021-999
 ; CURRENT APPLICATION NUMBER: US/10/320,797
 ; CURRENT FILING DATE: 2002-12-16
 ; PRIOR APPLICATION NUMBER: 60/341,261
 ; PRIOR FILING DATE: 2001-12-17
 ; NUMBER OF SEQ ID NOS: 3361
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 2305
 ; LENGTH: 2466
 ; TYPE: DNA
 ; ORGANISM: Cryptococcus neoformans
 US-10-320-797-2305

Alignment Scores:

Pred. No.: 8,926-229 Length: 2466
 Score: 2301.50 Matches: 435
 Percent Similarity: 73.0% Conservative: 107
 Best Local Similarity: 58.63% Mismatches: 155
 Query Match: 61.19% Indels: 45
 DB: 16 Gaps: 6

US-10-028-384-6 (1-718) x US-10-320-797-2305 (1-2466)

Qy 13 GlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaIleSerSerArg 32
 Db 223 GAAAGCTTTGCGCTTCATCATCTCGCGTTAATATGCGGCTCGGATTCGAAAGTGA 282
 Qy 33 LeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspProThrPheAsnTyr 52
 Db 283 TTGTTTGGGTGATCAGATTCGATCTGTCATCCAGAAATTCGCCCTGGTTCAACTAC 342
 Qy 53 ArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnTrpPheAsp 72
 Db 343 CGAGCCTCGAAAGTCTTTGTTAAACAGGGTTTCTACGAGTCTGGAAGTCTGACCCC 402
 Qy 73 ArgThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThr 92
 Db 403 TCCGCTTGGTACCTCTCGGAGAACTCTCGGTACCACTCTATCTCGCTTGTATGTC 462
 Qy 93 ThrSerAlaPheIleThrPheAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArg 112
 Db 463 AGCTCTGGACGTGATTGTCATGCTCTTCGGGCA---ATCAATATGCGCGTGGACATTGCG 519
 Qy 113 AsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGlu 132
 Db 520 AATGCTGTGCTCTCTTGCACCTGGATTTCTGGATTGATCTGCTGGCGACTTATCTT 579
 Qy 133 PheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaGlyPheIleAlaIle 152
 Db 580 TTCACCACTGAAATGCTTACACCACTACAGTGGTCTATTGGCGCGCTTTCATGGCAAT 639

Qy 153 ValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIle 172
 Db 640 GTACCCGGGATACATCTCTCGATCTCTCGCGGTTCTTATGACACAGCAAGCATTTGCCATC 699
 Qy 173 ThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMet 192
 Db 700 TTCCTCTTGATGAGCTCTCTACTCTTGGAATTAAGCCCTCAAAACCGGTAGCTCATTT 759
 Qy 193 HisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrVal 212
 Db 760 TGGGTATGATCAGCTGCTCTGTTCTACGGTGGATGTTGCTGATGGGTGGTTCAGTT 819
 Qy 213 PheIleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSer 232
 Db 820 TTCATCACCAACATGATCCATTCACGCGCTTCTTCATTCATGCGGCGGTTCAC 879
 Qy 233 SerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGln 252
 Db 880 AACCGCTTTTATACCGCTTACTCTTCCTGATGTCATTCGAACTATTCGCTCCATGCAG 939
 Qy 253 IleProPheValIcLysPheLeuProIleArgSerAsnAspHisMetAlaLeuGlyVal 272
 Db 940 GTCCCTTTGTGGAGTCTCTCCCATCCGAACCTCTGAGCAGATGGCGGCTTGGGTGT 999
 Qy 273 PheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAla 292
 Db 1000 TTCGGTCTTGATCAGCTGATCGGATTCGGAAGTGTGCGAGCTCTGCTGCGTGGCAAG 1059
 Qy 293 LysPheLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyVal 312
 Db 1060 CAATTCAGCTCTCTCTCAAGCTTTTGTGTCGCGCTTATTCCTGCTCAGTTTGTGTC 1119
 Qy 313 LeuSerAlaLeuThrTyrMetGlyLeuIleAlaProThrPheGlyArgPheTyrSerLeu 332
 Db 1120 CTGTCACCTTGATCTCTCTGGATGATCGCCCTTCTGCTGGAAGATTTTATCTCTT 1179
 Qy 333 TrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGln 352
 Db 1180 TGGGATCTGCTGATCGGAAGTCCACATGCCATTTATTCCTCGCTCCGAACACCAAG 1239
 Qy 353 ProValSerTrpProAlaPhePheAspPheThrHisPheLeuIleTrpLeuPheProAla 372
 Db 1240 CCACCGCTTGGCGCTCAATCTACTTTCGACTCGAAATGCTTATCTCTTTTCCCTGCC 1299
 Qy 373 GlyValPheLeuLeuPheLeuAspLysAspGluHisValPheValIleAlaTyrSer 392
 Db 1300 GGTGCTCTTCTGCTGTTTCAAGGAGCTTCGCGATGAGCAGATCTTCATCATCTTATGCC 1359
 Qy 393 ValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProVal 412
 Db 1360 GTTCTCAGTGCCTATTTTGGCGGTGTCATGTTGCTGCTTATGCTGTCATCAGCTGTT 1419
 Qy 413 IleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeuAsp----- 430
 Db 1420 GTCTGTGTTTCTCGCCATTCGCTTCTCCAACTTCTCGAGGCGTATATGACCCGCTC 1479
 Qy 431 -----PheLysThrSer 434
 Db 1480 ATCCCGAAAGCGACGAGGAGCTGGCGAGTCTCAGACGCGAGGTGTCCTCAAGTCCAAAG 1539
 Qy 435 AspArgLysTyrAlaIleLysProAla-----AlaLeuLeu 446
 Db 1540 GCGAAGAAGATGGCGCTGCGCAACGCCAATAAGACGCGGTCTCTTTCACAGGATATTG 1599
 Qy 447 AlaLeuLeuIleValSerGlySerPhe----- 455
 Db 1600 AGCGGCAAGTGTCTCGGCACTTTTGGTCTCGACACTCGATTTCGTGGTTCCTCAT 1659
 Qy 456 ---IlePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyr 474
 Db 1660 CTCTCTGCT 1719

Db 777 GATTGCCATCTTCTGCTGGTGGTTCCACATCTTCTTATGATCAAGGCTGTCAAAAATGG 836
 Qy 189 ySerileMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpCl 209
 Db 837 GTCATCATGTTGGGAGCGCTGACCGCACATATCTACGGCTACATGGTTCGGCATGGG 896
 Qy 209 yGlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMetG1 229
 Db 897 TGGGTATGCTCTTATACGAACCTGATCCCCCTGACGATTTTGTCTCTTCTGTCATGGG 956
 Qy 229 yArgTyrSerSerLysLeuTyrSerAlaTyrThrTyrThrTyrAlaIleGlyThrValAl 249
 Db 957 TAGATACAGCACTCGCATACATAGCTATACCATGATGATGCGCTGGGACATTTGGC 1016
 Qy 249 aserMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAl 269
 Db 1017 TAGCATGCAGATTCCTTCGTCGGATTTTTCCTATCCGAACACAGCAGCACATGTCGC 1076
 Qy 269 aleu-----G 271
 Db 1077 CTGGGTAGGCTTGTGCTTTCGCAAGTTCAAATGGCGCTTCAGAGCTGACATTTTGCAG 1136
 Qy 271 lyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerT 291
 Db 1137 GTGCTTCGGCTGCTTCAGCTTCGCTTCGCGAGTTTGTCCGACGCTTCGTTCCAA 1196
 Qy 291 hrAlaLysPheLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyValV 311
 Db 1197 GCAACAGTTCCAGAGACTTCGACCGCCGATGATTCATCACCTTCGGTCTCGGTTTCG 1256
 Qy 311 alGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTyrThrGlyArgPheTyrS 331
 Db 1257 TTGGGCTAGTTGTTCTGACTGTGACGGAGTGATGCTCTTCGAGCGCGCATCTACT 1316
 Qy 331 erLeuTrpAsnThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluH 351
 Db 1317 CCTTGTGGACACGTGCTATCCAAAATCCACATTCCTCATCTGCTCAGTCTCGAAC 1376
 Qy 351 isGlnProValSerTrpProAlaPhePhePheAspThrHisPheLeuIleTrpLeuPheP 371
 Db 1377 ACCAGCCACCGCTTGGCAGCGTCTCTTCGATCTGAATCTGTAATCTGCTGCTGCTTCC 1436
 Qy 371 roAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaT 391
 Db 1437 CGGAGGTGCTACATAGCTTCGCTGACCTCAGACGACGATGCTTCGTCATATCT 1496
 Qy 391 yrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrp 411
 Db 1497 ACTCGGCTCTTCGAGCTACTTCGCGGTGTTATGGTCGACTAATGCTGACCTTGACCC 1556
 Qy 411 roValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu--- 429
 Db 1557 CTATTGTGTGTGCGCTGCTCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1616
 Qy 430 -----AspPheLysThrSerAsp----- 435
 Db 1617 CTACCCCTCCGACACCGCTCTGAAGCGAAACGAATGAAGACTCGTCTTCACAACCTC 1676
 Qy 436 -----ArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValS 452
 Db 1677 TTCGCTCAGTTCGGAGCCCAATGTGGAATCACCTCCCATGTTCTTAGATATAGTGA 1736
 Qy 452 erGlySerPheIlePheTyrLeuTyrLeuPheValPheHisSerThrTyrValThrArgT 472
 Db 1737 CGGCTCTGTGTGCTGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796
 Qy 472 hrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaL 492
 Db 1797 ATGCATATCTTCTTCCTTCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1856
 Qy 492 euIleAspAspPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerL 512
 Db 1857 TCATTGAGCATTATCGTGAGGCTTACTACTGCTTCTGCTGCTGCTGCTGCTGCTGCTG 1916

Qy 512 yValAlaAlaAlaTrpTrpAspTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrL 532
 Db 1917 AATCATGTCATGTTGGGANTATGGTATCAATTCGTTGGTGGATCGCGACCGCCACCT 1976
 Qy 532 euValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaS 552
 Db 1977 TGGTTGACACACACACCTGGAACACACCATATTTGCTACGTTTGGTAAGCGCATGAGCT 2036
 Qy 552 erProGlnGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValI 572
 Db 2037 CAGCGAGAGAGTCACTACCCATCTCCCGCAGCATGATGTCATTAAGTCTGCTGGTGG 2096
 Qy 572 lePheGlyLeuLeuGlyPheGlyAspAspIleAsnLysPheLeuTrpMetIleA 592
 Db 2097 TGTTCGTGCTGCTGCTAGGTTATTCTGCGCATGACATTAACAAATCTTATGATGATGCTC 2156
 Qy 592 xGleSerGluGlyIleTrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluG 612
 Db 2157 GTATCCCGAAGTATCTGCGCCGATGAGGTTAAAGCGGCGACTTCTTACTGACGCG 2216
 Qy 612 lyGluTyrArgValAspAlaAlaArgAlaSerGluThrMetArgAsnSerLeuLeu--- 629
 Db 2217 GTGAATATGCTGTCAGCATGAGGAGCGACCCCACTATGCGCAACAGCTTGTATGATGAT 2276
 Qy 630 -----Tyr 630
 Db 2277 TCCTCTCTCTCTCTTATCTACTGCGTGCATCAGCTAAATTCACAAAATCTAGGTAT 2336
 Qy 631 LysMetSerTyrLysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
 Db 2337 AAAATGCTCTTATACAAATTCAACTCTCTCTGCGTGGGCCAAGCTTGTGACCGGCTC 2396
 Qy 651 ArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyrPheAspGluValPhe 670
 Db 2397 CGTGGGTCAAACTT---CCACAGAGCGCTCAGCTCTCTACCTCAGAGAGCTTTC 2453
 Qy 671 ThrSerGluAsnTrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGlyArg 690
 Db 2454 ACAGCGAGAACTGGATCATCTGATCTACAGGTCAAGGATCTTGCAACCTTGGCCGA 2513
 Qy 691 ThrLeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgAspSerIleLys 710
 Db 2514 GACCAACACCGCTGTTGCTTCACAAA---GGTCTCAAGAAAAAGCGGATACAAAG 2570
 Qy 711 Arg 711
 Db 2571 AGG 2573
 RESULT 8
 US-10-128-714-139
 ; Sequence 139, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Weng
 ; APPLICANT: Tishkoff, Daniel
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Eroshkin, Alexey M
 ; APPLICANT: Lemieux, Sebastien M
 ; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
 ; TITLE OF INVENTION: Methods of Use
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT FILING DATE: 2002-04-23
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR FILING DATE: 2001-07-09

```

; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 139
; LENGTH: 3969
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
; US-10-128-714-139

Alignment Scores:
Pred. No.: 1,32e-224 Length: 3969
Score: 2264.50 Matches: 461
Percent Similarity: 66.31% Conservativity: 98
Best Local Similarity: 54.69% Mismatches: 135
Query Match: 60.21% Indels: 150
DB: 15 Gaps: 11

US-10-028-384-6 (1-718) x US-10-128-714-139 (1-3969)
QY 15 IleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPhe 34
DB 698 CTTCTCGGATATATCTCTGTACCATTCGACGAGGAGGAGTGCACGCTGCTC 757
QY 35 AlaValIleLys----- 38
DB 758 AGCGTTATCCG-TAAGTCATCGAGGGAATCAAGTTGCTAAATTTGTTAATCTATCATCA 816
QY 39 -----PheGluSerIleIleHisGlu----- 45
DB 817 AATTGACTGATCATTTCTTTCTACAGGCTTCGAGAGTATCATCCAGGATGTAAGTAT 876
QY 45 ----- 45
DB 877 AGTCAATTGTATPACCTACGGCTTCGGCCTTTAAGGACCTCGCTTGGAGAAATTGATA 936
QY 46 -----PheAspProTrpPheAsnTyrArgAlaThrLysTyrLeuVal 59
DB 937 CTGATGCTCGGTATATAGTTGACCGGTGTTCACTTCCGAGCAACAAATATACTTAGTA 996
QY 60 AsnAsnSerPheTyrLysPheLeuAsnTrpPheAsp----- 72
DB 997 CAGAAATGTTTCTATAGTTTGGGATTTGGGATTTGGTATGATGACCGTATGTTGGTTCCCA 1056
QY 72 ----- 72
DB 1057 CAAGAGTGATGAAGCAACAGGTCACTGCTCGGAGTGTCTTAGGCTAATCGGCTCC 1116
QY 73 -----Arg-ThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLe 90
DB 1117 CAACATAGGAACATGGACCTCTGAGCGTGTACCGGTGCGAGTTATATCCCGTCT 1176
QY 90 uMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAs 110
DB 1177 CATGGTGACGAGCGGTGATCTACCATATCTTGGA--TTGCTTACTATCCCGCTGA 1233
QY 110 PileArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThr 130
DB 1234 TATTGCAACATCTCGTCTACTGCGCCAGATCTCCGCGCTGACTGCTGATTTGGCAAT 1293
QY 130 rTyrGluPheThrLysGluIleLysAspAla---SerAlaGlyLeuLeuAlaAlaGlyPhe 149
DB 1294 GTACTTGTGACATCGGATGTCCTTCGTCATCTGCGCTCTTCTTGCAGCAGCTTT 1353
QY 149 eIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAl 169
DB 1354 CATGGGAATCGCCCTGGTTTACATCTCCGATCAGTTGCTGAGCTACGATACGATACGAGC 1413
QY 169 aIleAlaIleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGl 189
DB 1414 GATTGCCATCTTCTGCTGTGTCTACATTTCTTATGATCAAGGCTGCAAAATAGG 1473
QY 189 ySerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGl 209
DB 1474 GTCTATCATGTGGGAGCGCTGACCGCACTATTCTACGGCTACATGGTCTCGCATGGG 1533
QY 209 yGlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGl 229
DB 1534 TGGGTATGTCTTCATTACGAACCTGATCCCGCTGCACGTTTTTGTCTTCTGTGATGG 1593
QY 229 yArgTyrSerSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAl 249
DB 1594 TAGATACAGCACTCGCATCTACATTAGCTATACCATGGTATGGCTGGGACTTTGGC 1653
QY 249 aSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAl 269
DB 1654 TAGCATGCAGATTCCTTCCTGTCGATTTTTCCTCTATCCGAACAGCAGCACCATGTCCG 1713
QY 269 aLeu-----G 271
DB 1714 CTTGGTAGGTCTCTGCTTCGCAAGTTCAATGGCGGTTCAGGAGCTGACATTTTGGNG 1773
QY 271 lYValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerT 291
DB 1774 GTGCTTCGGCTCTCTTCACTTGTGGCTTTCGCGAGTTTTCGCGAGCTTCTGTCCTCA 1833
QY 291 hrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeuGlyValV 311
DB 1834 GCAAGCAGTTCGAGAGCTTCTGACCGCATGATCTTCATCCTTCGCTCGGTTCGTTTCG 1893
QY 311 alGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrS 331
DB 1894 TTGGCTAGTTGTTCTGACTGTGACGGAGTGTCTCTCTGGAGCGCGCATCTTCTACT 1953
QY 331 erLeuTrpAspThrAsnTyrAlaIleHisIleProIleIleAlaSerValSerGluH 351
DB 1954 CTTGTGGACACTGGCTATGCCAAATCCACATTCCTCATCTGCTCAGTCTCGAAC 2013
QY 351 isGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPheP 371
DB 2014 ACCAGCCACCGCTTGGCAGCTTCTTCTTCGATCTGAACCTTCTCTGATCTGGCTTTCC 2073
QY 371 roAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAla 391
DB 2074 CGCAGGTGTCTATGTGTTCCGTGACCTCAAGACGAGCATGTCTTCGTCATTAICT 2133
QY 391 yrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrP 411
DB 2134 ACTCGTCTTTCGAGGTACTTCCCGCTGTATTATGTCGCACTAATGCTGACCTTGACCC 2193
QY 411 roValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu--- 429
DB 2194 CTATGTGTGTGTGCGCTGCTCTGCGCTGTCTGCTCAATCTCGACACCTATATGGCA 2253
QY 430 -----AspPheLysThrSerAsp----- 435
DB 2254 CTACCTCTCCGACACCGCTCTGAAGCGAAACGAATGAAGACTCGTCTTCACACCTTC 2313
QY 436 -----ArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLeuValS 452
DB 2314 TTCGCTCAGTTTCGGAAGCCCAATGTTGAATCACCTCCCATGTTCTTAAGATTATGTA 2373
QY 452 arGlySerPheIlePheTyrLeuPheValPheHisSerThrTrpValThrArgT 472
DB 2374 CGGCTCTGTGTCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2433
QY 472 hrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAla 492
DB 2434 ATGCATATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2493
QY 492 eIleAspAspPheArgGluAlaTyrTrpTrpLeuArgMetAsnSerAspGluAspSerL 512
DB 2494 TCATTGACGATATTCGTGAGCTTACTACTGGCTTCTGTCAGATATCTCTCTCTCTCTCT 2553
QY 512 ysValAlaAlaAlaTrpAspTyrGlyTyrGlnIleGlyMetAlaAspArgThrThrL 532

```

Db 2554 AAATCATGTCATGTTGGATATGGGTATCAATCGTGGCATGGCGGACCCCAACCT 2613
QY 532 euValAspAsnThrTrpAsnAsnThrHisleAlaValGlyLysAlaMetAla 552
Db 2614 TGGTGTACAAACACCTGGAACACACCATATCTAGGTTGGTAAAGGCGATGACT 2673
QY 552 erProGluGluLysSerTyrGluLysLeuLysGluHisAspValAspTyrValLeuVal 572
Db 2674 CACGCGAGAGTACAGTACCCATCTCCGCCAGCATGATGTCGATTACGGTCTGGTG 2733
QY 572 lePheGlyGlyLeuLeuGlyPheGlyAspAspLeuLysPheLeuTrpMetIleA 592
Db 2734 TGTTCGGTGGTCTCTAGGTTATCTCGCGCATGACATTAACAAATCTTTATGGATGGTCC 2793
QY 592 gIleSerGluGlyLeuTrpProGluLysGluLysGluArgTyrPheTyrThrAlaGluG 612
Db 2794 GTATCGCCGAGGATCTCGGCCGATGAGGTAAAGAGCGGACTTCTTTACTGCACGCG 2853
QY 612 yGluTyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeu 629
Db 2854 GTGAATATCGTGTGACGATGAGCGGACCCCACTATGCGCAACAGCTTGATGGAT 2913
QY 630 -----Tyr 630
Db 2914 TCCCTCTTCTCCCTCTATCTAGTGGTGCATCAGCTTAATTCACAAATCTAGGTAT 2973
QY 631 LysMetSerTyrLysAspPheProGluLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
Db 2974 AAAATGCTTTATTAACAATTTCAACTCTCTTCCCGTGGGCGCAAGCTGTGACCGCGTC 3033
QY 651 ArgGlnGlnMetIleThrProLeuAspValProLeuAspTyrPheAspGluValPhe 670
Db 3034 CGTGGGTCAAAACTT---CCACAGAGGCGCTCTCTACACTCGAAGAGCTTTC 3090
QY 671 ThrSerGluAsnTrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGlyArg 690
Db 3091 ACGAGCGAGACTGGATTCGATCTCTCAAGGTCAAGATCTTCAACACCTTGGCCGA 3150
QY 691 ThrLeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLys 710
Db 3151 GACCACACACAGGTGTGTCTTCGACAAA---GGTCTCAAGAAAAGCGGAGTCAAAAG 3207
QY 711 Arg 711
Db 3208 AGG 3210

RESULT 9
US-10-128-714-5139
; Sequence 5139, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroskin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-09-31
; NUMBER OF SEQ ID NOS: 8603

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5139
; LENGTH: 4603
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5139
Alignment Scores:
Pred. No.: 1-66e-224 Length: 4603
Score: 2264.50 Matches: 461
Percent Similarity: 66.31% Conservative: 98
Best Local Similarity: 54.69% Mismatches: 135
Query Match: 60.21% Indels: 150
DB: 15 Gaps: 11
US-10-028-384-6 (1-718) x US-10-128-714-5139 (1-4603)
QY 15 IleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPhe 34
Db 1061 CTTCTCGGATATTATCTCTGTGTACATTGTCAGCAGCAGCGATTGCCAGTCTGACTTTC 1120
QY 35 AlaValIleLys----- 38
Db 1121 AGGTTATCCG-TAAGTTCATCGAGGAATCAAGTTGCTAAATTTGTTAATCTATCATCA 1179
QY 39 -----PheGluSerIleIleHisGlu----- 45
Db 1180 AATTGACTGATCACATTTCTTTCTACCAGGCTTCAGAGATATCATCCAGGAATGAATAT 1239
QY 45 ----- 45
Db 1240 AGTCAATGTATATCTACGCGCTTCGGCCCTTTAACGACCTCGCTTGGACGAATTGATA 1299
QY 46 -----PheAspProTyrPheAsnTyrArgAlaThrLysTyrLeuVal 59
Db 1300 CTGATGCTCCGTGATATAGTTGACCCGTGGTTCACCTTCGAGCAACAAATATCTAGTA 1359
QY 60 AsnAsnSerPheTyrLysPheLeuAsnTrpPheAspAsp----- 72
Db 1360 CAGAAATGGTTTCTATAGCTTTTGGGATTTGTTGATGACCGTATGTGGTTCCCCACCGCA 1419
QY 72 ----- 72
Db 1420 CAAGATGTCATGAAGCAACAGGTACCTGTCTCTGAGTGTCTTAGCTAATCGGCGTCC 1479
QY 73 -----Arg-ThrTyrTrpProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLe 90
Db 1480 CAACATAGGAACATGTCATCTCTGGAGCGTGTCCCGGTGGCAGCTATATCCCGTCT 1539
QY 90 uMetThrThrSerAlaPheIleTyrHisAlaLeuArgAsnTrpLeuGlyLeuProIleAs 110
Db 1540 CATGGTGACGAGCGGCGTGTATCTACCATATCTTGGGA---TTCCTTACTATCCCGTCTGA 1596
QY 110 pIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAla 130
Db 1597 TATTCGCAACATCTCGCTCTACTGCGCCAGGATTTCCGCGCTGACTGCTCATGGCAAT 1656
QY 130 rTyrGluPheThrLysGluIleLysAspAla---SerAlaGlyLeuLeuAlaGlyP 149
Db 1657 GTACTTGTGTGATCCGAGATGTCTCTTCGCCATCTGCAGGTCTTCTTGCAGCAGCTTT 1716
QY 149 eIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAl 159
Db 1717 CATGGGAATCGGCCCTGTTTACATCTCCCGATCAGTTGCTGGAGAGCTTACCATACGAAGC 1776
QY 169 aIleAlaIleThrLeuLeuMetValThrPheMetPheTyrIleLysAlaGlnLysThrGl 189
Db 1777 GATTGCACTCTTCTGCTTGTGTTCCACATTTCTTCTATGATCAAGGCTGTCAAAATGG 1836
QY 189 ySerIleMetHisAlaThrCysAlaLeuPheTyrPheTyrMetValSerAlaTrpGl 209
Db 1837 GTCTATCATGTGGGAGCGGCTGACCGCCTATTCTACGGCTACATGTGTGCGCATGGG 1896

QY 209 yGlyTyrValPheIleThrAsnLeuIleProLeuHisValPhePheLeuLeuIleLeuMetGl 229
Db 1897 TGGGTATGTTTCAATACGAACCTGATCCCTGACGCTTTTGTCTTCTGTGATGGG 1956
QY 229 yAtgTyrSerSerLysLeuTyrSerAlaTyrThrThrTyrAlaIleGlyThrValAl 249
Db 1957 TAGNATCAGCAGCTCGCATCTACATTAAGCTATACCATGGTATGGCTGGGATTTGGC 2016
QY 249 aSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAl 269
Db 2017 TAGATGACAGATTCCCTGCTGCGATTTTGGCTATCCGAACACAGCAGCATGTCCGC 2076
QY 269 aLeu-----G 271
Db 2077 CTTGGGTAGTCTCTGCTTCGCAAGTTCAAATGSCCGTTCAGGAGCTGACATTTTCGAG 2136
QY 271 lyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheVallyGlyGlnIleSerT 291
Db 2137 GTGCTTCGGCCCTGCTTACGTTGTGGCTTCGCGAGTTTGTGCGAGCTTCGTTCCAA 2196
QY 291 hrAlaLysPheLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyValV 311
Db 2197 GCAAGCAGTTCAGAGACTTCTGACCGCCATGATCTTCATCACCCTTCGCTTCGCTTTC 2256
QY 311 alGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProThrThrClYArgPheTyrS 331
Db 2257 TTGGCTAGTGTGCTGACTGTGACCGGAGTGTGCTTCCCTTGGAGCGCGGATTTACT 2316
QY 331 erLeuThrAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluH 351
Db 2317 CTTGTGGGACACTGGCTATGCGAAATCCACATCCCATCATTCGCTTCAGTTCGGAAC 2376
QY 351 isGlnProValSerTyrProAlaPhePheAspThrHisPheLeuIleThrLeuPheP 371
Db 2377 ACAGCCACCGCTTGGCAGGCTTCTTCCGATCTGACCTTCCTGATCTGGCTTTC 2436
QY 371 rolaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaI 391
Db 2437 CGGAGGTGTCTACATGTGCTTCGCTGACCTCAAGGACGAGCATGTCTTCGCTCATCT 2496
QY 391 yrSerValLeuCySsrTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThp 411
Db 2497 ACTCGTCTTGGCAGTCTTCTGCGGGTGTATGCTCGCATATGCTGACCTTGACCC 2556
QY 411 roValIleCySsrValSerAlaValAlaLeuSerLysIlePheAspIleTyrLeu---- 429
Db 2557 CTATTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2616
QY 430 -----AspPheLysThrSerAsp----- 435
Db 2617 CTACCTCCGACACAGGCTGTGAAGCGAAACGATGAGACCTGCTCTTCACAACTC 2676
QY 436 -----ArgLysTyrAlaIleLysProAlaLeuAlaLeuAlaLysLeuIleValS 452
Db 2677 TTCGCTCAGTTCGGAAGCCCAATGTGGAAATCACCTCCCATGTTTCTAAGATTAAGTGA 2736
QY 452 erGlySerPheIlePheTyrLeuTyrLeuPheValPheHisSerThrThrPValThrArgT 472
Db 2737 CGGCGTCTGTGCTACCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2796
QY 472 hrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAla 492
Db 2797 ATGCATACCT 2856
QY 492 euIleAspAspPheAtgGluAlaTyrTyrTyrLeuArgMetAsnSerAspGluAspSerL 512
Db 2857 TCATTGACGATTATGCTGAGCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2916
QY 512 ysValAlaAlaTyrTyrAspTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThrL 532
Db 2917 AAATCATGTGATGCTGGATTTGGGTATCAAAATCGGTGGATGCGGACCGCCCAACCT 2976
QY 532 euValAspAsnAsnThrThrPAsnAsnThrHisIleAlaIleValGlyLysAlaMetAla 552

Db 2977 TGGTGTGACAAACACACCTTGGAAACACCCATATTGTCACGGTGTGTAAGCGCATGAGCT 3036
QY 552 erProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValI 572
Db 3037 CACGCGAGAGTTCAGTACCCCATCTCCGCGAGCATGATGTCGATTACGCTGCTGGTGG 3096
QY 572 lePheGlyGlyLeuIleGlyPheGlyGlyAspIleLeuLysPheLeuTyrMetIleA 592
Db 3097 TGTTCGGTGGTCTCTAGGTATTCTGCGCATGACATAACAAATCTTATGATGGTCC 3156
QY 592 gIleSerGluGlyIleTyrProGluGluIleLysGluArgTyrPheTyrThrAlaGluG 612
Db 3157 GTATCGCGAAGGTATCTGGCCGATGAGGTAAAGAGGGGACTTCTTTACTGCACGCG 3216
QY 612 lyGluTyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeu----- 629
Db 3217 GTGAATATCGTGTGAGATGGAGCGACCCCACTATGGCAACAGCTTGATGTATGAT 3276
QY 630 -----Tyr 630
Db 3277 TCCCTCTCTCCCTCTTATCTACTGCGCTGCATCAGCTTAATTCACAAATCTAGGTAT 3336
QY 631 lysMetSerTyrLysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
Db 3337 AAATGTCTTATTACATTTCACTCTCTTCCGTGCGGCGCAAGCTGTCCACCGCGTC 3396
QY 651 ArgGlnGlnMetIleThrProLeuAspValProLeuAspTyrPheAspGluValPhe 670
Db 3397 CGTGGTCAAACTT---CCACAGAAAGGCTCTCTCTACACTCGAAGAAGCTTTC 3453
QY 671 ThrSerGluAsnTyrMetValArgIleTyrGlnLeuLysAspAspAlaGlnGlyArg 690
Db 3454 AGGACGAGAACTGGATCTGATCTGATCTCAAGTCAAGATCTTCAACCTTGCCCGA 3513
QY 691 ThrLeuArgAspValGlyLeuLeuThrArgSerSerThrLysThrArgSerIleLys 710
Db 3514 GACCACACAGGCTGTGCTTCGACAAA---GGTCTCAAGAAAAAGCGGAGTACAAAG 3570
QY 711 Arg 711
Db 3571 AGG 3573
RESULT 10
US-10-128-714-2139
; Sequence 2139, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wendi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Ershkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2139
; LENGTH: 1848

! TYPE: DNA
! ORGANISM: Aspergillus fumigatus
US-10-128-714-2139

Alignment Scores:
Pred. No.: 6 01e-207 Length: 1848
Score: 2090.00 Matches: 394
Percent Similarity: 79.53% Conservative: 80
Best Local Similarity: 66.11% Mismatches: 93
Query Match: 55.57% Indels: 29
DB: 15 Gaps: 5

US-10-028-384-6 (1-718) x US-10-128-714-2139 (1-1848)

QY 63 PheTyrLysPheLeuAsnTrpPheAsp-----Arg-ThrTr 75
DB 6 TTCTATAGCTTTGGGATGGTTTGATGACCGCTAATCGCGCTCCCAACATAGGAACATG 65
QY 75 pTyrProLeuGlyArgValThrGlyThrLeuTyrProGlyLeuMetThrSerAl 95
DB 66 GCATCCTCTGGACGTGTACCGGTGGCAGCTTATATCCCGCTCTCATGTGACGAGCG 125
QY 95 aPheLeuTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAsnValGly 115
DB 126 CGTGATCTACCATATCTTGGCA---TTCTTACTATCCCGCTCGATATCGCAACATCG 182
QY 115 sValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThr 135
DB 183 CGTCTACTGGCGCGCAGGATCTCCGCGCTGACATGGCAATGCTACTTGTGACATC 242
QY 135 sGluLeuLysAspAla---SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaValPr 154
DB 243 CGAGATGCTCTCTCGCATCTGAGGTCTCTTGCAGAGCTTTCATGGATCGCC 302
QY 154 oGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleThr 174
DB 303 TGGTTACATCTCCCGATCAGTTGCTGAAGCTAGCATACGAACGATTCCTCTTCT 362
QY 174 uLeuMetValThrPheMetPheTyrLysAlaGlnLysThrGlySerIleMetHisAl 194
DB 363 GTTGTGTTCATCTTCTATGATCAAGCTGTCAAAATGGGTCTATCATGTGGGG 422
QY 194 aThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIl 214
DB 423 AGCGTGTACCGGACATATCTACGGCTACATGGTGTGCGCATGGGGTGGGTATGCTTCAT 482
QY 214 eThrAsnLeuLeuProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSer 234
DB 483 TACGAACCTGATCCCGCTGACGCTTTTGTCTCTGTCATGGGTAGATACAGCACTCG 542
QY 234 sLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIlePr 254
DB 543 CATCTACATATTAGTATACCATGATGATGGCTGGGACTTTGGCTAGCATCGCATTC 602
QY 254 oPheValGlyPheLeuProIleAArgSerAsnAspHisMetAlaAlaLeuGlyValPheG 274
DB 603 CTTGTGCGATTTTGTCTATCCGAAACAGCAGCCACATGTCCGCTTGGGTCTTCGG 662
QY 274 yLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLysPh 294
DB 663 CTTGTCTAGCTTGTGGCGAGTTGTCCGAGTTGTCCGAGCTTCTCCCAAGCAAGCATG 722
QY 294 eLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSe 314
DB 723 CCAGAGACTCTTGACCGGCGATGATCTTATCCATCCCTCGGTCTCGGTGCTGGGTAGT 782
QY 314 rAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpAs 334
DB 783 TGTTCGACTGTGACGGAGATGATGCTCTTGGAGCGCGGATTTCTCTCTCTTGGGA 842
QY 334 pThrAsnTyrAlaLysIleHisIleProIleAlaSerValSerGluHisGlnProVa 354
DB 843 CACTGGCTATGCCAAATCCACATTCCTCATCTGCTGATCGGAAACAGCCGAC 902

QY 354 lSerTrpProAlaPhePhePheAspThrHisPheLeuIleTrpLeuPheProAlaGlyVa 374
DB 903 CGCTTGGCGAGCGTTCTTCTTCGATCTGAACCTCTGATCTGGCTTTCCCGCAGGTGT 962
QY 374 lPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyrSerValle 394
DB 963 CTACATGTCTTCCGTGACCTCAAGGAGCAGCATGCTTCTGATTTACTTCTCTCTCT 1022
QY 394 uCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValleCy 414
DB 1023 TGGAGCTACTTCCGCGGTGTATGTCGCGACTAATGTCGACTTACCTGACCCCTATTGTG 1082
QY 414 sValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu----- 429
DB 1083 TGTTCGCGTGTCTGCGGTGTCTGTCATCTCGACCTATATGCGGACTACCTCC 1142
QY 430 -----AspPheLysThrSerAsp----- 435
DB 1143 GACACCGAGCGTCTGAAGCGAAACGATGAAGACTCTCTTCAACACTCTTCGCTCAGT 1202
QY 436 -ArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPh 455
DB 1203 TCGGAAGCCCAATGTGGAATCACCTCCATGTTTCTAAGATTATAGTACGCGCTCTGT 1262
QY 455 ellePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSe 475
DB 1263 TGTCTGCTACCTGCTCTCTGTTTGTGGCACTGACCTGGGTACATCGAATGCTACTC 1322
QY 475 rSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAs 495
DB 1323 TTCTCTCTCGGTGTCTGGCTAGTCCGATCGCTGACGGAAGCCCAATACATCATTCACGA 1382
QY 495 pPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAl 515
DB 1383 TTATCTGAGGTACTACTGCTTCTGCAAGATCTCTCAGAACGCCAAATCATGTC 1442
QY 515 aTrpTrpAspTyrGlyTyrGlnIleGlyLysAlaAspArgThrThrLeuValAspAs 535
DB 1443 ATGTTGGATTATGGTATCAATCGGTGGCTGATCGCGGACCCGCAACCTTGTGTGACAA 1502
QY 535 nAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGlu 555
DB 1503 CAACACCTGGAAACACCCCATATTGCTACGGTGGTAAAGCGATGATGCTACGCGGAGA 1562
QY 555 uLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyCl 575
DB 1563 AGTCAGCTACCCCACTCTCCCGCAGCATGATGCTAGTCTGCTGGTGTGCTGGTGG 1622
QY 575 yLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGl 595
DB 1623 TCTGTAGTATTATCTGGCGATGACATTAACAAATCTTATGATGCTCGTATCGCCGA 1682
QY 595 uGlyIleTrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrAr 615
DB 1683 AGTATCTGGCGGATGAGTTAAAGAGCGGACTTCTTTACTGCACCGCGTGAATATCG 1742
QY 615 gValLeuAlaAsgAlaSerGluThrMetArgAsnSerLeuLeuTyr 630
DB 1743 TGTTCAGCATGAGGAGCCCACTATCGCGCAACAGTTGATGATAT 1788

RESULT 11

US-10-128-714-1139
Sequence 1139, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Weng
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Broshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and

; TITLE OF INVENTION: Methods of Use
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US/10/128,714
 ; PRIOR FILING DATE: 2002-04-23
 ; PRIOR APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR APPLICATION NUMBER: US 60/287,066
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1139
 ; LENGTH: 1969
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus
 ; US-10-128-714-1139

Alignment Scores:

Pred. No.: 6,59e-202 Length: 1969
 Score: 2042.00 Matches: 394
 Percent Similarity: 74.41% Conservative: 80
 Best Local Similarity: 61.85% Mismatches: 93
 Query Match: 54.29% Indels: 70
 DB: 15 Gaps: 6

US-10-028-384-6 (1-718) x US-10-128-714-1139 (1-1969)

Qy	63	Phenylalanylserine	-----	72
Db	6	TTCTATAGCTTTTGGGATTGGTTGATGACCGTATGTGGTTCGCCACGCAAGAGTGC	-----	65
Qy	73	-----	-----	73
Db	66	ATGAAGCAACAGGTCACTGCTCGGAGTGTCTTAGGCTAATCGCGCTCCCAACATAGG	-----	125
Qy	74	ThrtPyrProLeuGlyArgValThrGlyThrLeuTyrProGlyLeuMetThr	-----	93
Db	126	AACATGGCATCTCGGGACGTGTACCGGTGGACGTATATCCCGGTCTCATGGTGC	-----	185
Qy	93	rSerAlaPheIleThrPheAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAs	-----	113
Db	186	GAGCGGCTCATCTACCATATCTTGGCA--TTCTTACTATCCCGCTCGATATTCGAA	-----	242
Qy	113	nValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTrpGluPh	-----	133
Db	243	CATCTGCTCTACTCGCGCAGGATCTCCGGCTGACTGCAATGGCAATGTACTTGGCT	-----	302
Qy	133	eThrLysGluIleLysAspAla---SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaI	-----	152
Db	303	GACATCGAGATGTCCTTCGCAATCTCGAGTCTTCTTGCAGCAGCTTTCATGGGAAT	-----	362
Qy	152	eValProGlyTyrIleSerArgSerValAlaGlySerTyrAspGluAlaIleAlaI	-----	172
Db	363	CGCCCTGGTTTACATCTCCGATCTGTCGAGTACGATACGAGGATGGCAT	-----	422
Qy	172	eThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMe	-----	192
Db	423	CTTCTGCTGTGTTCACATCTTCTTATGGATCAAGGCTGTCAAAATGGGTCTATCAT	-----	482
Qy	192	HisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrVa	-----	212
Db	483	GTGGGAGCCCTGACCGCACTATTCTACGGCTACATGGTTCGGCATGGGTGGGTATGT	-----	542
Qy	212	IleIleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMetGlyArgTyrSe	-----	232
Db	543	CTTCATTACGAACCTGATCCCTGACCGTTTCTCTCTGTCATGGGTAGATACAG	-----	602
Qy	232	rSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetG	-----	252

Db	603	CACTCGCATCTACATTAGCTATACCATGGTATGGCTGGGACTTTGGCTAGCATGCA	-----	662
Qy	252	nIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaLeu----	-----	270
Db	663	GATTCCCTTCGTGGATTTTGGCTATCCGAACAGGACCATATGTCGCTTGGGTAG	-----	722
Qy	271	-----	-----	274
Db	723	GTCCCTTGTCTTTCGCAAGTTCAAATGGCGCTTCAGGAGCTGACATTTTGCAGGTG	-----	782
Qy	274	IyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLysP	-----	294
Db	793	GCCTGCTTCAGCTTGTGGCTTCGCCAGTTCGCGAGCTTCGCGAGCTTCGTCACGA	-----	842
Qy	294	heLysValIleMetMetValSerLeuPheLeuLeuValLeuGlyValValGlyLeuS	-----	314
Db	843	TCCAGAGACTTCTGACCGCCATGATCTTCATCCTTCGCTTCGCTTCGCTTGGCTAG	-----	902
Qy	314	erAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpA	-----	334
Db	903	TTGTTCGACTGTGACGGGAGTATCGCTCTTGGAGCGCGCATTTCTACTCTCTGGG	-----	962
Qy	334	spThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProV	-----	354
Db	963	ACACTGGCTATGCCAAATCCCATCCCATCTTCGCTCAGTCTCGGACACACGCCCA	-----	1022
Qy	354	alSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPheProAlaGlyV	-----	374
Db	1023	CGCTTGGCCAGCGTCTCTTCGATCTGAATCTCTGATCTGCTTTCGCGCAGGTG	-----	1082
Qy	374	alPheLeuLeuPheLeuAspLeuLysAspGlnHisValPheValIleAlaTyrSerVal	-----	394
Db	1083	TTCTATCATGTCTCCGTGACCTCAAGACGAGCATGTCTTCGTCATATCTACTCGGT	-----	1142
Qy	394	euCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValIleC	-----	414
Db	1143	TTGCGAGCTACTTCGCGGTGTTATGTTCGACTTAATGTGCTGACCTTATGACCCCT	-----	1202
Qy	414	ysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu-----	-----	429
Db	1203	GGTTCGCGCTGCTCTGCGCTGTCTGCTCACTATATGCGGACATACCCCTCC	-----	1262
Qy	430	-----	-----	435
Db	1263	CGACACAGCGCTCTGAAGCAAAACGAATGAAGACTCGTCTTCCAAACCTTTCGCTCAG	-----	1322
Qy	436	--ArgLysTyrAlaIleLysProAlaLeuLeuAlaLysLeuIleValSerGlySerP	-----	455
Db	1323	TTGGAAGCCCAATGTTGGAAATCACCTCCCATGTTTCTAAGATTATAGACGGGCTG	-----	1382
Qy	455	heIlePheTyrLeuTyrPheValPheHisSerThrTrpValThrAlaTyrS	-----	475
Db	1383	TGTGCTACTGCTGCTCTGTTTGTTCGCACTGCACCTGGGTATACATGCAATGCACT	-----	1442
Qy	475	erSerProSerValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspA	-----	495
Db	1443	CTTCTCCCTTCGTTGTTTGGTATGCTGATGCTGCGGAGCCCAATACATTCATTGAC	-----	1502
Qy	495	ppPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaA	-----	515
Db	1503	ATTATCGTGGGCTTACTTACTGCTGCTGATATCTCTCAGAAATCAATCATGT	-----	1562
Qy	515	laTTrpTrpAspTyrGlyTyrGlnIleGlyMetAlaAspArgThrThrLeuValAspA	-----	535
Db	1563	CATGTTGGATATGGGTATCAATCGTGGCATGCGGACCCGCCCACTTGGTTGACA	-----	1622
Qy	535	enAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluG	-----	555
Db	1623	ACAACACCTCGAACAACACCATATGCTACGTTGGTAAAGCGCATGAGCTCACGGAGG	-----	1682
Qy	555	luLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyG	-----	575

Db 1683 AAGTCAGTACCCCATCTCCGCCAGCATGATGTCGATTAGGTGCTGGTGGTTCGGTG 1742
 Qy 575 lyleuileedlyPheGlyAspAspIleAsnLysPheLeuTrpMetIleAglleSerG 595
 Db 1743 GTCTGCTAGGTATTCTCGGCATGACATTAACAATTTCTTATGGATGGTCCGTATCGCG 1802
 Qy 595 lueGlyTrpProGluGluLeuLysGluArgTyPheTyTrpAlaGluGlyGluTyA 615
 Db 1803 AAGGTATCTGGCCCATGAGTTAAAGACGGGACTTCCTTACTGACCGGGTGAATATC 1862
 Qy 615 rGValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyR 630
 Db 1863 GTGTCGACGATGGAGCGACCCCACTATGCGAACAGCTTGATGAT 1909

RESULT 12

US-10-320-797-1305

; Sequence 1305, Application US/10320797

; Publication No. US20040014955A1

; GENERAL INFORMATION:

; APPLICANT: Broskin, Alexey M.

; APPLICANT: Zamudio, Carlos

; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND

; TITLE OF INVENTION: METHODS OF USE

; FILE REFERENCE: 10182-021-999

; CURRENT APPLICATION NUMBER: US/10/320,797

; CURRENT FILING DATE: 2002-12-16

; PRIOR APPLICATION NUMBER: 60/341,261

; PRIOR FILING DATE: 2001-12-17

; NUMBER OF SEQ ID NOS: 3361

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1305

; LENGTH: 2882

; TYPE: DNA

; ORGANISM: Cryptococcus neoformans

US-10-320-797-1305

Alignment Scores:

Pred. No.: 1.16e-200 Length: 2882
 Score: 2032.50 Matches: 433
 Percent Similarity: 60.86% Conservative: 105
 Best Local Similarity: 48.98% Mismatches: 159
 Query Match: 54.04% Indels: 191
 DB: 16 Gaps: 14

US-10-028-384-6 (1-718) x US-10-320-797-1305 (1-2882)

Qy 13 GlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaIleSerSerArg 32
 Db 223 GAAAGCTTGTTCGGCTTCATCATCTCGCGTTAATATGCGGTGCTCGGATTGGAATCGA 282
 Qy 33 LeuPheAlaValIleLysPheGluSerIleIleHisGluPhe----- 46
 Db 283 TTGTTTGGCGTATCAGATTGGAATTCGATCTGTCATCCACGAATTGTGAGTCAAACTATCCGC 342
 Qy 47 -----AspProTrpPheAsnTyR-- 52
 Db 343 AGGATTCCTCAATGCTAAGCTGTGTGTCCTTCATTTCAGTGACCCCTCGGTCAACTA-GT 401
 Qy 53 -----ArgAlaThrIleTyR 57
 Db 402 AAGGCTATTCTCGAGGTCAAACTTCCTACTGATTTATTCCTAGCGAGCCCTCGAAAGT 461
 Qy 57 rIeuValAsnAsnSerPheTyLysPheLeuAsnTrpPheAspArgThrTrpTyR 77
 Db 462 TCTTTTAAACAGGGTTCTACGAGTTCTCGAATCTGTTGACCCCTCGCGTGTGTATCCC 521
 Qy 77 oleuGlyArgValThrGlyGlyThrLeuTyTrpGlyLeuMetThrThrSerAlaPheI 97
 Db 522 TCTCGGACAGCTGTCGTCACCAAGCTCTATCTCCCTGCTGATGTCACGCTCTGGACTGAT 581
 Qy 97 eTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysVal 117
 Db 582 TTGGCATGCTCTTCGGGCA---ATCAATATGCGCGTGACATTCGCAATGTCTGTGTCT 638

Qy 117 uPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyR----- 131
 Db 639 CTTTGACCTGGATTTTCTGGATTGACTCCCTGGCGGCACTTATCTGTCACTGTAACAAAT 698
 Qy 132 -----GluPheThrLysGluIleLysAs 139
 Db 699 TATCTGATCATTCGCATACACATAACATGAGAGCGTTGTAGTTTACCACTGAAATGCTAC 758
 Qy 139 pAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyRLeSerAr 159
 Db 759 ACCATCAGCTGCTTATTGGCGCGCTTTCATTGGCATTGTACCCGATACATCTCTCG 818
 Qy 159 gSerValAlaGlySerTyRAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPh 179
 Db 819 ATCTGTGGCGGTTCTTATGACACAGAGCCATTGCCATCTCTCTTGTGATGAGCTCCT 878
 Qy 179 eMetPheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAla 199
 Db 879 CTACTCTTGGATTAAGCGCGTCAAAACCGGTAGCTCAATTTGGGGTATGATCACTGCCCT 938
 Qy 199 uPheTyRTrpMetValSerAlaTrpGlyGlyTyRValPheIleThrAsnLeu----- 217
 Db 939 GTTCTACGGGTGATGGTGTGTCATGGGTGGTTACGTTTTCATCACCACCAACAG-TATGT 997
 Qy 218 -----IleProLeuHisValPh 223
 Db 998 CGCTCGGCCCTCAATTGAATGTCTGTTTACTCTTTTGCAGTGATTCATTTGCACGCCCT 1057
 Qy 223 eLeuLeuIleLeuMetGlyA-gTyRSerSerLysLeuTyR-SerAlaTyRThrTrpTy 243
 Db 1058 TGTTCATTGATGATGGCGGTTCAACACCGGCTTTATACCGCTTACTCTTCTCCGTTA 1117
 Qy 243 rAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleArgse 263
 Db 1118 TGTCAATGAACTATCGCTCCATGACGAGTCCCTTTGTGGAGTCTCTCCCAATCCGAAC 1177
 Qy 263 rAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAs 283
 Db 1178 CTCGTAGCACATGGCGGCTTGGGTGTTTTCGGTCTGTGTACAGTGATCGGATTCGTGCA 1237
 Qy 283 pPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuPh 303
 Db 1238 AGTGTCCGACGACTGCTGCTGGCAGCAATTCAGCTCTCTCTCMAAGCTTTTGTGCT 1297
 Qy 303 eLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyRMetGlyLeuIleAl 323
 Db 1298 GGCCTGATTCCTGCTCAGTTTGTGCTCCCTCGTCACTTTCATCTCTCTGATGATCGC 1357
 Qy 323 aProTrpThrGlyArgPheTyRSerLeuTrpAspThrAsnTyRAlaLysIle----- 340
 Db 1358 CCCCCTCGCTGGAAGATTTTATCTCTTTGGGATCTGGCTATGCGAGGTCCACAGTGA 1417
 Qy 341 -----His-IleProIleIleAla 347
 Db 1418 GTCAAAATGCCATACCTTCGGGATCTATATTTATTTGTAACATAGTCCCATTTATGCGCT 1477
 Qy 347 eValSerGluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeu 367
 Db 1478 CGGTCTCCGAACACACGACCCCGCTGGCCCTCATTTCTACTTGTACCTCGAATGCTTA 1537
 Qy 367 leTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValP 387
 Db 1538 TCTTCTTTTCCCTCGCGGTGCTCTTGTGNGTTCCTCAAGGAGCTTCGCGATGAGACATCT 1597
 Qy 387 heValIleAlaTyRSerValLeuCysSerTyRTrpPheAlaGlyValMetValArgLeuMet 407
 Db 1598 TCATCATCATTTATGCCGTTCCTAGTGCCTATTTTTCGGGTGTCATGTTTCGACTTATGC 1657
 Qy 407 eThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspI 427
 Db 1658 TTGTCAACGCGCTGTGTGTGTGTTTCTCCCGCATTTGCGTCTCCCAACTTCTCGAGG 1717

QY	427	leTyLeuAsp-	-----	430
Db	1718	CGTATATTGACCCCGTCATCCCAGAAAGCAGCAGGAAGCTGGCAGTCTCAGACGCAGG	1777	
QY	431	-----PheLysThrSerAspArgIlystyrAlaIleLysProAla	-----	443
Db	1778	TGTCTCCAAGTCCAAGGCCAAGAAGATGGCGCTGCCAACGCCCAATAAGAGCGGTTCCT	1837	
QY	444	-----AlaLeuLeuAlaLysIleValSerGlyserPhe	-----	455
Db	1838	CTTTCACAGGTATTTTGAGCGGCAAGTGTCTCGGCATCTTTGGTCTCGACACTCGAT	1897	
QY	456	-----IlePheTyrlEuTyrlLeuPheValPheHisSerThrrpv	469	
Db	1898	TTGCTGTGGTTTCCATCTCTCTCTCTCTCATCTTTGCTCTCACTGCACATATG	1957	
QY	469	alThrAtgThrAlaTyrlSerSerProSerValValLeuProSerClnThrProspGlyL	489	
Db	1958	TGACTTCAACAGCGTATCTTCGCTTCAGTGGTACTTCGATCGGCAACC CGATGGTA	2017	
QY	489	ysLeuAlaLeuIleAspAspPheArgGlualaTyrlTrpLeuArgMetAsnSerAspG	509	
Db	2018	GCCAAAATATCATGTATGATTTCCGAGAGGCTTACTCTGGATTTCGCAAAACACCCGC	2077	
QY	509	IuAspSerLyIsValAlaAlaTrpTrpAspTy-clTyrlcInileGlyMetAlaAspA	529	
Db	2078	AAGACAGCGCTCATCATGCTCTGTGGGATTTACGGCTACCGATCCCTGGTAGGCTGATC	2137	
QY	529	rgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysA	549	
Db	2138	GCCCCACCTTGTGTGATAACAATACCTTGGAAATAACACCCACATTCGCCACAGTTGGTAGG	2197	
QY	549	laMetAlaSerProGluGluLysSerTyrlGluIleLeuLysGluHiAspValAspTyrv	569	
Db	2198	CCATGGCTTCCACAGAGATGTGCGATATCTTCTTGAGGAGCATGATGATGATTCAG	2257	
QY	569	alLeuValIlePheGlyGlyLeuIleGlyPheGlyIyAspAspileAsnLysPheLeut	589	
Db	2258	TTCTTGTGATCTTTGGGGGCTTATGGGCTACTCTGGTACCATATCAACAAGTTTTTGT	2317	
QY	589	rpMetIleArgIleserGluGlyIleTrpProGluIleLysGluArgtyrPheTyrt	609	
Db	2318	GGATGGTTAGGATCTCACAGGTGAATGGCTGACGAGGTGCGAGAGTCAACTTCTTTA	2377	
QY	609	hrAlaGluGlyGlyTyrlArgValasAlaArgAlaserGlu	622	
Db	2378	CTCAAAGAGGGAGTAGTGTCTGATCAGAGGCGTG-CGTCTATTTTGTATTATTGCC	2436	
QY	623	-----ThrMetArgAsnSerLeuLeuTyrl	631	
Db	2437	TGGGAAATTTGTCACAATCTGGCGAGCACCCCTACTATGAAGAACTCTCTCATGTACA	2496	
QY	631	ysMetSerTyrlLys	635	
Db	2497	AAATGTCCTTACTA-CCGGTAGGCAAAATACATTTGCGATAGATGAACATCGTCTGCAT	2555	
QY	636	-----AspPheProGlnLeuPheAsnGlyGlyGluAlaThrAspArgValArgGlnGlnM	654	
Db	2556	ATCTCAGCTTCCCGAGCTTATGGTGGACACCCGCTCAAGACAGGGTTCGAGGCCAAA	2615	
QY	654	etIleThrProLeuAspValProProLeuAsp-TyrPhe	666	
Db	2616	TTATCCCTCCFAACAGTGT--ACTCTGTACTCTTGGTAGTTTAGGTTTAGGTTTGAAGAT	2672	
QY	667	-----AspGluValPheThrSerGluAsn	674	
Db	2673	AAGTGTATAAGACGCTAATGTGCTCACAATTTCTGCGACAGAGCGTTCCATCCGAAAAAT	2732	
QY	675	TrpMetValArgIleTyrlGlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAsp	694	
Db	2733	TGGATCGTCAGGATCTCACAGGTCACAGAGGAGATCCCATTTGGACGAGACCAACGCC	2792	
QY	695	Val-----GlyGluLeuThrArgSerSerThrLysThrArgArgSerIle	709	

```

Db      2793 GTTACTGCTGAACAGGGGGTAAGAAGTTGAAGAAGAGTCCTAGTAGCCAGTGAGGGCGTG 2852
          |||:::   ::|||
QY      710 LysArg 711
          |||||
Db      2853 AAGCGG 2858

RESULT 13
US-10-320-797-305
; Sequence 305, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroskin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; PRIORITY FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 305
; LENGTH: 4738
; TYPE: DNA
; ORGANISM: Cryptococcus neoformans
US-10-320-797-305

Alignment Scores:
Pred. No.:      2.48e-200      Length:      4738
Score:           2032.50       Matches:     433
Percent Similarity: 60.86%    Conservative: 105
Best Local Similarity: 48.98%  Mismatches:   159
Query Match:      54.04%      Indels:      191
DB:                16         Gaps:         14

US-10-028-384-6 (1-718) x US-10-320-797-305 (1-4738)
QY      13 GlutThrLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArg 32
          :::::::|||||
Db      1079 GAACGTGTGTGCCTCATCCTCGCGCTTAATATGCGTGCTGCGNTTGGAAATCGA 1138
          ::::::::::::::::::::
QY      33 LeuPheAlaValIleLysPheGluSerIleIleHisGluPhe----- 46
          ::::::::::::::::::::
Db      1139 TTGTTTGGGTGATCAGATTCTGATCCAGAAATTGTGAGTCAAAATCTATCCGC 1198
          |||:::|||||
QY      47 -----AspProTrpPheAsnTy-- 52
          ::::::::::::::::::::
Db      1199 AGGATTCCTCAATGCTAACGTCTGTGCTCTTCATTTTCAGTGACCCCTGGTTCAACTA-GT 1257
          |||:::|||||
QY      53 -----ArgAlaThrLysTy 57
          |||:::|||||
Db      1258 AAGGCTATCTTCGAGGTCAAACCTTCCTACTGATTTATTCCTAGCCGAGCGCTCGAAAGT 1317
          |||:::|||||
QY      57 rLeuValAsnAsnSerPheTyLysPheLeuAsnTrpPheAspAspArgThrTrpTyPr 77
          |||:::|||||
Db      1318 TCTTGTAAACAAGGGTTTACGAGTTCGAACTGTTTTGACCCTCCCGCTTGTGTGCC 1377
          |||:::|||||
QY      77 cLeuGlyArgValThrGlyGlyThrLeuTyProGlyLeuMetThrThrSerAlaPheIl 97
          |||:::|||||
Db      1378 TCTCGGCAAGAACTGTCGGTACCAGCTCTATCTCGCTTGGTGGTACGCTCTGGACTGAT 1437
          |||:::|||||
QY      97 etrPhisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValle 117
          |||:::|||||
Db      1438 TTGGCATGCTCTTCGGGCA--ATCAATATGCCGTGGACATTCGCAATGTCTGTGCTCT 1494
          |||:::|||||
QY      117 upheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTy-- 131
          |||:::|||||
Db      1495 CCTTCACCTGGATTTTCTGGATTGACTGCTCGGGGACTTATCTGTCTAGTGTAACAAT 1554
          |||:::|||||
QY      132 -----GlupHeThrLysGluIleLysAS 139
          |||:::|||||

```

1555 TATCTGATCATTCATACATGAGAGCGTTGTAGTTTCCACCACTGAAATGCTAC 1614
Db
139 pAlaserAlaGlyLeuLeuAlaGlyPheIleAlaIleValProGlyTyrIleSerAr 159
Qy
1615 ACCATCAGCTGGTCTATTTGGCGCGCTTTCATTCGATGATGACCGGATACATCTCG 1674
Db
159 gSerValaGlySerTyrAspAenGluAlaIleAlaIleThrLeuLeuMetValThrPh 179
Qy
1675 ATCTGTGCGCGGTCTTATGACACGAGCAATGCGCATCTTCTCTTCATGAGCTCCT 1734
Db
179 eMerPheThrIleValAlaGlnIlySerThrGlySerIleMetHisAlaThrCysAlaAlaIle 199
Qy
1735 CTACTCTTGGATTAAAGCGGTCACAAACCGGTACTCATTTTGGGTGATGATCATCTGCTT 1794
Db
199 uPheTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeu----- 217
Qy
1795 GTTCTACGGTGGATGTTGTCATGGGTGGTTAGCTTTTACATCCACACAG-TATGT 1853
Db
218 -----IleProLeuHisValPh 223
Qy
1854 CGCTCGGCCCTCAATTGATGCTGTTTACTCTTTTGCAGTATTCATTCGACGCTT 1913
Db
223 eLeuLeuLeuMetGlyArgTyrSerSerIlySerLeuTyrSerAlaTyrThrTrpTyr 243
Qy
1914 TGTCTCATTTGTCATGGCGAGTTCACAAACCGGCTTTATACCGCTTACTCTCTCTGTA 1973
Db
243 rAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleArgSe 263
Qy
1974 TGTCAATGGAACTATCGCTCCATGACGAGTCCCTTTTGGAGTTCCTCCCATCCGAC 2033
Db
263 rAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAs 283
Qy
2034 CTCTGAGCACATGGCGGCTTGGGTGTTTTCGGTCTTTGACAGTCTGATCGGATCGTGA 2093
Db
283 pPheValIlyGlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuPh 303
Qy
2094 AGTGTCCGACACTGCTGCTGGCAGCAATTCAGCTCTTCTCAAGACTTTTGTGT 2153
Db
303 eLeuLeuValLeuGlyValGlyLeuSerAlaLeuThrTyrMetGlyLeuAl 323
Qy
2154 GCGCGTATTCCTGCTCAGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2213
Db
323 aprThrPheGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle----- 340
Qy
2214 CCCCTTCGCTGGAAGATTTATCTCTTTGGGATCTGGCTATGCGAAGTCCACAGTGA 2273
Db
341 -----His-IleProIleIleAlas 347
Qy
2274 GTCAAAATGCCATACCTTCGCGATCTATATTTATGCAACATAGTGCCTATTTGCT 2333
Db
347 erValSerGluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeu 367
Qy
2334 CCGTCTCCGAACACACGCCCTTGGCCCTCATCTTACTTTGACCTTCGAAATGCTTA 2393
Db
367 leTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuTyrAspGluHisValP 387
Qy
2394 TCTTCTTTTCCCTGCGCGGTCTCTCTGCTGCTTCAAGGAGCTTCGCGATGAGCAGTCT 2453
Db
387 heValIleAlaTyrSerValIleCysSerTyrPheAlaGlyValMetValArgLeuMetL 407
Qy
2454 TCATCATATTTATGCGGTCTCTGATGCTCTATTTTCCCGGTGTCATGCTGCTGCTGCT 2513
Db
407 euThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerIlyIlePheAspI 427
Qy
2514 TTGTATCACGCTGTTGCTGTTTCTCCGCAATGCTGCTTCTCCAAACTCTTCGAGG 2573
Db
427 leTyrLeuAsp----- 430
Qy
2574 CGTATTTGACCCCGTCTATCCCCGAAAGCAGACGAGAGCTGGCGAGTCTCAGACGCGAG 2633
Db
431 -----PheLysThrSerAspArgLysTyrAlaIleLysProLa----- 443
Qy
2634 TTGTCTCCAGTCCAAAGCGGAGGAAGATGGCGCTGCCAACGCCCAATAAGAGCGGTCT 2693
Db

444 -----AlaLeuLeuAlaLysLeuIleValSerGlySerPhe----- 455
Qy
2694 CTTTCACAGGTATTTGAGCGCAAGTCTGCTCGGCATCTTTGGTCTCGACACTCGAT 2753
Db
456 -----IlePheTyrLeuTyrLeuPheValPheHisSerThrTrpV 469
Qy
2754 TTGCTGTGGTTTCCATCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2813
Db
469 alThrArgThrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyL 489
Qy
2814 TGCATTCACACGGTATTTCTTCGCTTCAGTGTACTTGCATCGCAACCCCGATGTA 2873
Db
489 ysLeuAlaLeuIleAspAspPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspG 509
Qy
2874 GCCAAATATCATGATTTCCGAGAGGCTTACTTGTGATTCGCAACCAACCGCCG 2933
Db
509 luAspSerLysValAlaAlaTyrTrpAspTyrGlyTyrGlnIleGlyGlyMetAlaAspA 529
Qy
2934 AGACAGCGTCTCATGCTCTGCTGGGATTAAGCTACCGTACCGATCGTGTATGGCTATC 2993
Db
529 rgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysA 549
Qy
2994 GCCCACCTCTTGTATACCAATACCTGGAATAACACCCACATTCGCCACAGTTGGTAAG 3053
Db
549 laMetAlaSerProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrV 569
Qy
3054 CCATGGCTTCCAAAGAGTCTGCAATATCTTCTGGAAGCATGATGCTGATTCAG 3113
Db
569 alLeuValIlePheGlyGlyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuT 589
Qy
3114 TTCTTGTGATCTTTGGGGCTTATTTGGGCTACTCTGCTGACGATATCAACAAAGTTTGT 3173
Db
589 rMerIleArgIleSerGluGlyIleTrpProGluGluIleLysGluArgTyrPheTyr 609
Qy
3174 GGATGGTGGATCTCACAGGTGAATGGCTGACGAGGTGCAGAGTCACTACTACTTTA 3233
Db
609 hrAlaGluGlyIleValArgValAlaAlaSerGlu----- 622
Qy
3234 CTCAAAGAGGGAGTATCTGCTGATGACAGGGGTG-CGTCTATTTTGTATTATTGTC 3292
Db
623 -----ThrMetArgAsnSerLeuLeuTyrL 631
Qy
3293 TGGCGAAATTTGCTGACATCTGGGACACCCCTTACTATGAAGACTCTCTCATGTACA 3352
Db
631 ysMetSerTyrLys----- 635
Qy
3353 AAATGCTTACTA-CCGCTAGTCAAAATACATTGCTGATAGTGAACATCGTCTGACAT 3411
Db
636 -----AspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnM 654
Qy
3412 ATCTCAGCTTCCCGAGCTTTATGGTGGACACCCGCTCAAGACAGGTTTCGAGGCCAAA 3471
Db
654 eIleThrProLeuAspValProProLeuAsp-TyrPhe----- 666
Qy
3472 TTATCCCTCTCAACAGTGT---ACTCTTGATCTCTTGTGTAAGTTTAGGTTTGAAGAT 3528
Db
667 -----AspGluValPheThrSerGluAsn 674
Qy
3529 AAGTGTATAGACGCTAATGCTGCTCACATCTCTCGACAGCGAGGCTTCACATCCGAAAT 3588
Db
675 TrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAsp 694
Qy
3589 TGGATCGTCAGGATCTCAAGGTCAAGGAAGATCCCATTTGACGAGACCAAGGCC 3648
Db
695 Val-----GlyGluLeuThrArgSerSerThrLysThrArgArgSerIle 709
Qy
3649 GTTACTGCTCGAAACGGGGTAAAGAGTTGAAGAGTCTCTAGTCCAGTGGGGCGTG 3708
Db
710 LysArg 711
Qy
3709 AAGCGG 3714
Db

RESULT 14

US-10-028-384-7
 ; Sequence 7, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREAU, Kevin
 ; APPLICANT: MCBRIDE, Claude
 ; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
 ; FILE REFERENCE: 5600-74
 ; CURRENT APPLICATION NUMBER: US/10/028,384
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 7
 ; LENGTH: 2417
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: AF132552
 ; DATABASE ENTRY DATE: 1999-04-27
 ; RELEVANT RESIDUES: (1)..(2417)

US-10-028-384-7

Alignment Scores:

Pred. No.:	2,92e-200	Length:	2417
Score:	2027.50	Matches:	394
Percent Similarity:	68.22%	Conservative:	117
Best Local Similarity:	52.60%	Mismatches:	187
Query Match:	53.91%	Indels:	51
DB:	15	Gaps:	7

US-10-028-384-6 (1-718) x US-10-028-384-7 (1-2417)

QY	12	PheGlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaIleSerSer	31
DB	113	TACAGCAGCCAAATCAACCTTCGCGCATCTCTGTAATCGCTGGTGGCGGATTTCTCTCT	172
QY	32	ArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspProTrpPheAsn	51
DB	173	CGCCTCTTCGGCGGATCCGTTTCGAGTCGATATCTCATGAGTTGATCGGTGATCAAC	232
QY	52	TyrArgAlaThrLysTrpLeuValAsnSerPheTrpLysPheLeuAsnTrpPheAsp	71
DB	233	TACCGGGCCACCGCTACATGTTGCAGATGGTTGGTACAACTTCCTCACTGGTTCCGAC	292
QY	72	AspArgThrTrpTrpProLeuGlyArgValThrGlyThrLeuTrpProGlyLeuMet	91
DB	293	GAGCGCGCATGGTATCCGCTCGGCGAGGATTCGGCGGCTACCTCTATCCCGGCTGATG	352
QY	92	ThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIle	111
DB	353	ATTAGTCCGGCGGAATC--CATTTGGCTGTGCGACGTACTCAACATACGGTCCATATT	409
QY	112	ArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTrp	131
DB	410	CGTGACATCTCGGTTCCTCGCGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTAC	469
QY	132	GluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaGlyPheIleAla	151
DB	470	CTGCTGACCAAGGAGCTGTGTGTCGCGGCGCGGCTCTTCGCGCGGCTTCATCGCC	529
QY	152	IleValProGlyTrpIleSerArgSerValAlaGlySerTrpAspAsnIleAlaIleAla	171
DB	530	ATCGTGGCTGGCTACATCAGTAGTGGTGGCTGGATCGTACGATACAGGAGGCAATGCC	589
QY	172	IleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIle	191
DB	590	ATATTCGCGCTGCATTCACCTACTCTCTGTGGTGGCTCAGTGAGAGCTGATCCGCG	649
QY	192	MetHisAlaThrCysAlaAlaLeuPheTrpPheTrpMetValSerAlaTrpGlyTrp	211
DB	650	TTCTGGTCCGCGCGGCGGCTTTGTCCTACTTCTACATGGTGTCCGCTGGGTGGCTAC	709

QY	212	ValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMetGlyArgTrp	231
DB	710	GTGTTTCATCATCAACCTGATACCCCTGCGACGTCTTCGTACTGCTATTTATGGCGAGTAC	769
QY	232	SerSerLysLeuTrpSerAlaTrpThrTrpTrpAlaIleGlyThrValAlaSerMet	251
DB	770	TCGCGCGGTCTGTCAGCACCTTACAGCCTTCTACATCTCGGACTGCTGTCTCCAG	829
QY	252	GlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaLeuGly	271
DB	830	CAGATCCCTTCGTGGGATTCACACCGATACGACCATGCAACACATGCTGCGTGGGA	889
QY	272	ValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThr	291
DB	890	GTGTTTGTCTCTATGCGCGTGGCCACCTTCGCGCATTTGCAGTCCGCTGTGCGGC	949
QY	292	AlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeuGlyValVal	311
DB	950	AACGAGTTCGCGAAGCTGTTTCATCTGCGCGGATTCGTGCGGCTTGGCGCTTTGTG	1009
QY	312	GlyLeuSerAlaLeuThrTrpMetGlyLeuIleAlaProTrpThrGlyArgPheTrpSer	331
DB	1010	CCGCTGCTGCTCACCATGCTGGCGGTGTCGCGGCTGGAGTGGAGCGCTTCTACTCG	1069
QY	332	LeuTrpAspThrAsnTrpAlaLysIleHisIleProIleIleAlaSerValSerGluHis	351
DB	1070	CTGTGGGATACCTGCTACGCGCAAGATCCACATTCCTCCATCATTCGCTGCGAGCAT	1129
QY	352	GlnProValSerTrpProAlaPhePhePheAspThrHisPheIleTrpLeuPhePro	371
DB	1130	CAGCCACACCTTGGTCTCTCTTCTTGTGATCTGCATCTCTGCTGGCGCTTCCCA	1189
QY	372	AlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTrp	391
DB	1190	GTGGGAGTGGTACTGCATCAAGCAGATCAACACGAGCGCGCTTTTCGTGCTGTATC	1249
QY	392	SerValLeuCysSerTrpPheAlaGlyValMetValArgLeuMetLeuTrpLeuTrp	411
DB	1250	GCCATAGTGGTTCCTGCTGCTGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	1309
QY	412	ValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTrpLeuAspPhe	431
DB	1310	GTGCTGTCATCTGCGCGGAGTGGCTTTTCGGGACTGTGTGATGTGTCTCTGCAAGAG	1369
QY	432	LysThrSerAspArg	436
DB	1370	GATTCGCTAAGCGAATGGGCAAGCGCTGTACGCAAGGCTGGCAAGCTGATTAAG	1429
QY	437	-----LysTrpAlaIleLys	441
DB	1430	GATTCATTCGAGAAGACGCTGTACGCAAGGCTGGCAAGCTGATTAAG	1489
QY	442	ProAlaAlaLeuLeu-----AlaLysLeuIleValSerGlySer	454
DB	1490	CATGATCCCGCAGCAGATATCTGGCGTCACTCCAACTGAAAGATTAATGTTATTTGGCC	1549
QY	455	PheIlePheTrpLeuTrpLeuPheValPheHisSerThrTrpValThrArgThrAlaTrp	474
DB	1550	GTCTAATGCTGTGTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1609
QY	475	SerSerProSerValValLeuPro---SerGlnThrProAspGlyLysLeuAlaLeuIle	493
DB	1610	TCCAGTCCCTCCATGCTTGGCTTCCCAACACAGATGGATTCGCGCAACATTTTA	1669
QY	494	AspAspPheArgGluAlaTrpTrpLeuArgMetAsnSerAspGluAspSerLysVal	513
DB	1670	GACGATTTACAGAGGCTTACTACTGCTTTCGACAGACACTGCCGATGATGCTGCGGT	1729
QY	514	AlaAlaTrpTrpAspTrpGlyTrpGlnIleGlyMetAlaAspArgThrLeuAla	533
DB	1730	ATGCTTGGTGGGATACCGATACCGATACCGGGAATGCGAAGACGACGCTAGTG	1789

534	QY	AspAsnAsnThrTrrPAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerPro	553
1790	Db	GATATAATACGTCGAAACAATAGCTTACATAGCGCTGGTTGGCAAGGCAATGCTTCAACC	1849
554	QY	GlulGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePhe	573
1850	Db	GAGGAGAGTCCTACGAAATTATGACATCTCTGACCTGGACTAGCTTGGTGATCTTT	1909
574	QY	GlyGlyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuTrrPMetIleArgIle	593
1910	Db	GCGCGTGTGATCGCTCATTTCTGCGCATGATATCAACAAGTTCCTGTGGATGGTCCGAATT	1969
594	QY	SerGluGlyIleTrrPProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGlu	613
1970	Db	GCTGAGGAGAGCATCCCAAGGACATTAGGAAAGCGATTACTTTACCGACCGCGGTAA	2029
614	QY	TyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTrrLysMetSer	633
2030	Db	TTCAAGGGTAGATCGCAAGGTGTTCGGCCCTGCTCAACTGGCTTATGTACAAATTAAGC	2089
634	QY	TyrLysAspPheProGlnLeu-----PheAsnGlyGlyGlnAlaThrAspArgVal	650
2090	Db	TACTACATTCGGGGAATTGAAGTTGGACTACAGAGGTCCTATCTGGATATGATCGCAC	2149
651	QY	ArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyrPheAspGluValPhe	670
2150	Db	CGTAACGCCGTCATTTGGGANTAGGACTTC---GATCTGACCTACTCGGAGGAGCGCTAC	2206
671	QY	ThrSerGluAsnTrrPMetValArgIleTyrGlnLeuLysLysAsp-----	685
2207	Db	ACCACAGAACACTGCGCTTGTTCGCATCTATAGGTTGAAGAGCCGCATGAGTTCAATAGA	2266
686	QY	-----AspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArgSerSer	702
2267	Db	CCATCACTGAAGACACAGGAGAGAAAGATTCTCTCCAGCAACTTCATTTCGAGAAAGAAC	2326
703	QY	ThrLysThrArgArgSerIleLysArg	711
2327	Db	TCCTAAGCGTCGCAAGGGCTACATACGA	2353

RESULT 15

```

US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, Gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)..(2469)
US-10-028-384-3

```


Page 12

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:52:55 ; Search time 558.27 Seconds
(without alignments)
4707.175 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 3761

Sequence: 1 MGSDRSCVLSVFQTLKLVI.....TRSTKTRRSIKRPELGLRV 718

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131519_8030/app_query.fasta_1.3740
-DB=EST -QFMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=Dlosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN 1 1 13960 @runat_14122004_131519_8030 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLCCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2013.5	53.5	2709	AK018758	AK018758 Mus muscu
2	1998	53.1	4513	BC044321	BC044321 Xenopus 1
3	1977.5	52.6	2669	AK012153	AK012153 Mus muscu
4	1939	51.6	3802	AK030363	AK030363 Mus muscu
5	1932	51.4	2730	AK081547	AK081547 Mus muscu
6	1928.5	51.3	2656	BC028897	BC028897 Mus muscu
7	1919.5	51.0	3643	AK077877	AK077877 Mus muscu
8	1836	48.8	2118	AY418284	AY418284 Homo sapi
9	1812.5	48.2	2727	AK087470	AK087470 Mus muscu

10	1807.5	48.1	2103	9	AY418286	AY418286 Mus muscu
11	1689	44.9	2419	3	CNS0ADYZ	EX816490 Arabidops
12	1465	39.0	2079	9	AY418285	AY418285 Pan trogl
13	1430.5	38.0	1088	9	CNS06CLO	AL392680 T7 end of
14	1404.5	37.3	2334	3	AY310157	AY310157 Rattus no
15	1303.5	34.7	981	9	CNS06YOT	AL420435 T3 end of
16	1138	30.3	972	9	CNS07ABF	AL436261 T7 end of
17	1100.5	29.3	869	9	CNS06DY5	AL394419 T3 end of
18	1009.5	26.8	978	7	CR413615	CR413615
19	946.5	25.2	1141	6	CD505084	CD505084 CDA71-G02
20	932	24.8	504	7	CF821733	CF821733
21	929.5	24.7	1345	7	CR606556	CR606556 full-length
22	916	24.4	897	3	CO008636	CO008636 EST796971
23	911	24.2	836	7	CO012548	CO012548 EST800883
24	910.5	24.2	1019	5	EX404479	EX404479
25	896.5	23.8	960	6	CD793779	CD793779 EST65140
26	895	23.8	852	7	CNA58732	CNA58732 UI-M-HS0
27	894.5	23.8	876	6	CD794446	CD794446 EST65807
28	894.5	23.8	878	6	CD788585	CD788585 EST659946
29	894.5	23.8	891	6	CD778215	CD778215 EST649576
30	894.5	23.8	919	7	CR439276	CR439276
31	884.5	23.5	960	7	CN062150	CN062150 Ag2_P35_1
32	862	22.9	1145	5	EX446691	EX446691
33	856.5	22.8	863	7	CR440471	CR440471
34	852	22.7	933	5	BUS13073	BUS13073 AGENCOURT
35	852.5	22.7	862	5	EX713937	EX713937
36	852.5	22.7	863	7	CR442100	CR442100
37	852	22.7	1124	7	CK030365	CK030365 AGENCOURT
38	851	22.6	823	7	CF736482	CF736482 UI-M-HD0
39	848.5	22.6	863	5	EX741571	EX741571
40	845	22.5	836	5	CNS0172X	CNS0172X
41	829.5	22.1	903	6	CA487501	CA487501 Drosophl
42	826	22.0	931	7	CO008635	CO008635 EST796970
43	824	21.9	593	8	BZ295215	BZ295215 CG1357.r1
44	821	21.8	950	5	BUS15954	BUS15954 AGENCOURT
45	813.5	21.6	894	5	BUS16368	BUS16368 AGENCOURT

ALIGNMENTS

RESULT 1

AK018758

LOCUS

DEFINITION

AK018758 2709 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male liver cDNA, RIKEN full-length enriched
library, clone:130006C19 product:hypothetical Oligosaccharyl
transferase (OTase) STT3 subunit containing protein, full insert
sequence.

ACCESSION

AK018758

VERSION

AK018758.2 GI:26384577

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

2

Genome Res. 10 (10), 1617-1630 (2000)

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kitagawa, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

11042159

20499374

11042159

20499374

11042159

20499374

11042159

Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequence. *Genome Res.* 10 (11), 1757-1771 (2000) 20530913

11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection. *Nature* 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420, 563-573 (2002) 6 (bases 1 to 2709)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, H., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y. Direct Submission Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

On Dec 10, 2002 this sequence version replaced gi:12858635. Please visit our web site (<http://genome.gsc.riken.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCGCGCACTCGAGTTTCTTTTCTTNN 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOUR.

Location/Qualifiers

1..2709

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM_DB:1300006C19"

/db_xref="taxon:10090"

/clone="1300006C19"

/sex="male"

/tissue_type="liver"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

76..2547

/note="unnamed protein product; hypothetical oligosaccharyl transferase (OTFase) SRT3 subunit containing protein (InterPro|IPR003674, evidence: InterPro) Putative"

/codon_start=1

CDS

/protein_id="BAB31390.1"

/db_xref="GI:12858636"

/translation="MAEPSAPESKHKSLNSPMSGLMALNSRHHGHGPTQASRAA
AAPKPPAGSGGLSQPAGQSLLSFTILFALAGSRLFAVIRFESIHEFDPM
FNVRSTHLAGHGFVFLNWDERAWYPLGRVGTVPGLMTAGLTHWLTINTIT
VHROVCVFLAPTFESGLTSISTFLTRMLNQCGLAACFAIVPGVIRSVAQSF
NGGIALFALQFTYIYLWKSVKTSVFVTMCCLSYFVMSAWGVFIINLIPLHVFV
LMLLQYRSKRYIAYSTLIYVGLILSMQIPFVGPOPIRTSEHMAAGVFLALQVAF
QYLRDLTKQEFQTLFFGLVSLAGAVLSVLYITGTIAPFASAGVFLVLTSAIV
HIPITASVSEHQFTTWVFFDLHLVTCFAGLMFCIKINDERSVFLVLTSAIV
AGVRLMLTLPVVCMLSAFNSVFEHLGDDMKRNPVDSDDDDKRNPNGLY
DKAGVRKHVTQEPKPEGLGNIKSIIVTLMMLMLMFAVCTSSSNAYSSSVL
ASVNDGNIILDDPREAVFLRONTDEHARVMSWDYGYOIAQMANRTTILVNDNTN
NSHIALVQKAMSSNETAAYKIMRSLDVLVLEGGVLYGSDIDKFLVMVRAEGE
HPKDLREGDYFTQQGSEFVFKAGSPTLNCMLMKSYRFGEMOLDPFTPGPQRTN
AIGNKDKIFKLEAFTSEHVLRIYIKVAPDNRETLCHFRVTNIVPKQKLSKKT
TKRKGYVKNKLVPFKKKTSTTV"

2693..2698

/not_a="putative"

2709

/note="putative"

polyA_signal

polyA_site

ORIGIN

Alignment Scores:

Pred. No.:	2,04e-213	Length:	2709
Score:	2013.50	Matches:	399
Percent Similarity:	69.05%	Conservative:	112
Best Local Similarity:	53.92%	Mismatches:	177
Query Match:	53.54%	Indels:	53
DB:	3	Gaps:	12

US-10-028-384-6 (1-718) x AK018758 (1-2709)

Qy 12 PheGlnThrIleuLysLeuValIlePheValAlaIlePheValAlaIleSerSer 31

Db 268 TGGCAGCGTGTCTTCCTTCACCACTCTCTCGCTGGCTGGCGCTTCAGCTCG 327

Qy 32 ArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspProThrPheAsn 51

Db 328 CGCTCTTCGCGTCTATCCGCTTCGAGAGCATCATCCAGAGTTCGACCGCTGGTTTAC 387

Qy 52 TyrArgAlaThrIleLysLeuValAlaAsnSerPheTyrLysPheLeuAsnTrpPheAsp 71

Db 388 TATAGATCAACACATCATCTTCATGGATTCTATGAGTTCTTAATTTGGTTTGTAT 447

Qy 72 AspArgThrTrpTyrProLeuGlyArgValThrGlyThrLeuTyrProGlyLeuMet 91

Db 448 GAAAGAGCATGGTACCCACCTGGGAGAGATAGTGGTGGCACCCTTTACCCAGGTTGATG 507

Qy 92 ThrThrSerAlaPheIleThrHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIle 111

Db 508 ATAACAGCTGGCTTATTT---CATTTGGATTTTAAATACATTTGAACATACAGTTTCATA 564

Qy 112 ArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyr 131

Db 565 AGAGATGTGTGTATCTCTTCGACCACTTTTAGCGGCTTACATCATCATCATCTTCAGTTC 624

Qy 132 GluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAla 151

Db 625 CTGTAACTAGAGAACTGTGGAAACCAAGAGCAGGAGCTTCTAGTCTGCTTCATTCGTCT 684

Qy 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171

Db 685 ATCGTACAGGATACATATCTCGTTCAGTGGCGGATCTCTTGTATATGAAGCAATTCGCC 744

Qy 172 IleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIle 191

Db 745 ATTTTGGGCTTCAGTTCACTTACTTACTTATGGTAAAGTCTGTGAGACCGGCTCTGTG 804

Qy 192 MethHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyGlyTyr 211

Db 905 TTCTGGACAACTGCTGCTGCTTGTCTTGTATTTCTTACATGCTCTGCTGGGGAGGTTAT 864

Db 2366 -----ACTAGTCAAACTCGAGTCAACCA-----CATCTGCCCA 2403
 Qy 713 u 713 :
 Db 2404 A 2404

RESULT 4
 AK030363
 LOCUS
 DEFINITION Mus musculus adult male pituitary gland cDNA, RIKEN full-length enriched library, clone:5330402C17 product:OLIGOSACCHARYL TRANSFERASE SRT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus], full insert sequence.

ACCESSION AK030363
 VERSION AK030363.1
 KEYWORDS GL:26326354
 SOURCE HTC; CAP trapper;
 ORGANISM Mus musculus (house mouse)

REFERENCE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 Carninci, P., and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 2 9279253
 PUBMED 10349636

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Kiteunai, T., Tashiro, H., Itoh, M., Konno, H., Akiyama, J., Nishi, K., Hazama, M., Nishine, T., Hara, A., Sumi, N., Ishii, Y., Nakamura, S., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, O., Okazaki, Y., Muramatsu, M., Inoue, K., Kira, A., and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 695-699 (2001)
 5

6 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 3802)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, K., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takada, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
 Direct Submission

JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
 COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 FEATURES
 Location/Qualifiers
 1..3802
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM DB:5330402C17"
 /db_xref="taxon:10090"
 /clone="5330402C17"
 /sex="male"
 /tissue_type="pituitary gland"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 133..2250
 /note="unnamed protein product; OLIGOSACCHARYL TRANSFERASE SRT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus] (SWISSPROT|P46978, evidence: FASTA, 99.7%ID, 100%length, match=2115) putative"
 /codon_start=1
 /protein_id="BAC26921.1"
 /db_xref="GI:26326355"
 /translation="MTKGLFRLSYERKQDTLLKLLLSMAAVLSFSLFAVLRPESV IHEPDFYNTTRFLAEFGYFHNFDRAWYPLGIIGTIVGLMTGAAYIVH LHPFTTDIRNVCFPLAPFSSTFTVTYVHLTKLADAGLAAATVAVPVYISR SVAGSYDNEGIAIFCMLLTYVMKIKAVTGSIVMAAKCALAVYVSSMGVYGLNL IPIHLVLMGTGRSHRYVAYCTVCLGTSLSMOISFVGFQFVLSSEHMAAFVGL QTHAFVDYICSKLNFQCFELPVSIVSLGVFLLVGLALLMTKISFTWGRFVSLLL DQSIHAFNIIASVSEHQPTWSSYIFDLQLLVFPFVGLVYFNSLSDARIFIMY PKNVMSFVAVMLMDLVLAPVWGLSGIGVSVQVSVYMKLDISRDKSKKQDSTY IPIKNEVASGMILVMAFLITYFTSTWTSYVSSPSIVLSARGDGSRIIPDDREA YYLHNTNPEDAKVMWDYGYOITANAKTILVDNTWNTTHSRVGMAMASTBEKA YEIMRELDVSVLVIFGLTGYSSDDINKFLMWVRCGSGTETGRHKENDYVTPGFE RVDEGSPVLLNGLMKVCYVRGQVYVTEAKRPPGDRVRNAEIGNKDFELVDLEAY TTEHWLVRIYKVDLNRGLSR"

ORIGIN
 Alignment Scores:
 Pred. No.: 7,61e-205 Length: 3802
 Score: 1939.00 Matches: 378
 Percent Similarity: 69.64% Conservative: 106
 Best Local Similarity: 54.39% Mismatches: 183
 Query Match: 51.56% Indels: 28
 Gaps: 10
 DB: 3

US-10-028-384-6 (1-718) x AK030363 (1-3802)

Qy 14 ThrilleuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeu 33
 Db 178 ACACCTTCAAGCTTCATCCTCGATGGCTGCTGTATCTTTTCTACTCGCTT 237
 Qy 34 PheAlaValIleLysPheGluSerIleIleHieGluPheAspProTrpPheAsnTrpArg 53
 Db 238 TTTCGTGCTGAGATTGAAGTTCATCCATCCATGAGTTTGATCCGCTTATTAATTCGG 297
 Qy 54 AlaThrLysTrpLeuValAsnSerPheTrpLysPheLeuAsnTrpPheAspArg 73
 Db 298 ACTACCGGTTTCGCTGAGGAGGGGTTTATAATTCCATACTGTTTGTGATCCCG 357


```

99279253
MEDLINE
PUBMED
REFERENCE
AUTHORS
Carninci, P., Shibata Y., Hayatsu, N., Sugahara Y., Shibata K.,
Itoh M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalisation and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED
REFERENCE
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2730)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koyas, S., Kuwihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
Location/Qualifiers
1..2730
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C130038C18"
/db_xref="taxon:10090"
/clone="C130038C18"
/tissue type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev stage="16 days embryo"
125..2243
/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
(B5) (INTRAGAL MEMBRANE PROTEIN 1) homolog (Mus musculus)
(SWISSPROT)P46978, evidence: FASTA, 99.7%ID, 100%length,
match=2115)
putative"
2717..2722
/note="putative"
2730
polyA_site
misc_feature
polyA_signal
polyA_site
ORIGIN
Alignment Scores:
Pred. No.: 2,75e-204 Length: 2730
Score: 1932.00 Matches: 378
Percent: 69.68% Conservative: 107
Best Local Similarity: 51.31% Mismatches: 182
Query Match: 51.37% Indels: 29
DB: 3 Gaps: 10
US-10-028-384-6 (1-715) x AK081547 (1-2730)
QY 14 ThrileLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeu 33
Db 170 ACACCTTCTAAAGCTTCTCATCTCGATCGCTGCTGTGTATCTTTCTCTACTGCTT 229
QY 34 PheAlaValIleLysPheGluSerIleIleHisGluPheAspProtrpPheAsnTyArg 53
Db 230 TTTCGTGCTGAGATTGAAAGTGTCATCATGAGTTGATCGTACCTTAATATCGG 289
QY 54 AlaThrLysTyLeuValAsnSerPheTyLysPheLeuAsnTrpPheAspArg 73
Db 290 ACTACCGGTTCTGGCTGAGAGGGGTTTATAAATTCATTAACCTGGTTGATGACCGG 349
QY 74 ThrTptTyProLeuGlyArgValThrGlyGlyThrLeuTyProGlyLeuMetThrThr 93
Db 350 GCTTGTTACCTTGGGCGAATCATTTGGAGGAACAATTTACCCAGGTTTATGATCACT 409
QY 94 SerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProleAspIleArgAsn 113
Db 410 TCTGCTGCAATCTACCATGACTC---CATTTCTCATATCACCATTGACATTCGGAAT 466
QY 114 ValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyGluPhe 133
Db 467 GTCGTGTTTCTCTGGCGCGCTTTCTCTCTTACCACCATCGTTACGACCACTT 526
QY 134 ThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleVal 153
Db 527 ACCAAGAGCTCAAGGATGAGGAGCTGGGCTTCTGCTGCTCCATGATGCTGAGTT 586
QY 154 ProGlyTyriLeSerArgSerValAlaGlySerTyAspAsnGluAlaIleAlaIleThr 173
Db 587 CTTGGGTATATTTCTCGATCTGAGTCTGCTCTCTATGATTAATGAAGAAATTTGATCTTT 646
QY 174 LeuLeuMetValThrPheMetPheTrpIleLysValGlnLysThrGlySerIleMetHis 193
Db 647 TCGATGCTGCTTACTTACTACATGTGATCAGGAGTGAAGACTGGTTCATCTATTGG 706
QY 194 AlaThrCysAlaAlaLeuPheTyPheTyMetValSerAlaTrpGlyGlyTyValPhe 213
Db 707 GCTGCCAAGTGTGCTCTGCTTATTTCTATGATGCTCTTCTTATGAGGAGGATGTTGTTTC 766
QY 214 IleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMetGlyArgTySerSer 233
Db 767 CTGATCAACTTGTCTCTTACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 826
QY 234 LysLeuTySerAlaTyThrThrTyAlaIleGlyThrValAlaSerMetGlnIle 253
Db 827 CGGATCTACGTAGCTACTGTACTGTTTACTGCTGGGACCACTTCTTCTATGAGAT 886
QY 254 PropheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGlyValPhe 273

```


Db 887 TCCTTTGGTTTCCAGCCGCTCTTTCATCAGAACACATGCGACGCTTTGAGGTGTT 946
 Qy 274 GlyLeuileGlnIleValAlaPheGlyAspPheVallyGlyGlnIleSerThrAlaTy 293
 Db 947 GGTCTCTGTGCAGTCCATGCTTCTGTAGATTACCTGCGCAGCAAGTGAATCCACAGCAA 1006
 Qy 294 PheLysValIleMet-----MetValSerLeu-----PheLeuileLeuValLeuGly 309
 Db 1007 TTCGAAAGTCTTTTCGGAGGTGTATCTCCCTGTTGCTTGTGCTTCTCTCCTCAGTGGGA 1066
 Qy 310 ValValGlyLeuSerAlaLeuThrTyrMetGlyLeuileAlaProThrThrGlyArgPh 329
 Db 1067 GCTCTC-----CTCATGCTACAGGAAATTTCTCCCTGCGACAGGGGGT 1114
 Qy 329 eTySerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSe 349
 Db 1115 CTACTCTCTGCTGATCCCTCTTAATGCTAAGAAATACATTCCTCATTTGCTCTGTTTC 1174
 Qy 349 rGluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeuileTrpIe 369
 Db 1175 TGAGCACCAGCCCAACACCTGCTCTCTCTACTATTCTTTCATCAGCTCTCTGCTTCAT 1234
 Qy 369 uPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValI 389
 Db 1235 GTTTCAGTGGCTCTATTACTGCTTTAGCAACCTCTCTGATGCTCGGATTTTATCAT 1294
 Qy 389 eAlaTySerValLeuCySerTyrPheAlaGlyValMetValArgLeuMetLeuThrIe 409
 Db 1295 CATGTATGTTGTGACACACATGTTCTTTTCAGCTGTAAATGGTGGCTAATGCTGTAT 1354
 Qy 409 uThrProValIleCyValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyr-- 428
 Db 1355 GGCACCTGTATGTGCTTCTTTCTGGCATGGTGTTCCTCCAGGTGCTGCCACATATAT 1414
 Qy 429 -----LeuAspPheLysThrSerAspArgLys-----ATGGCTTTCT 1525
 Db 1415 GAAATCTGGACATAAGTCCGCCACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1474
 Qy 438 rAlaIleLysProAlaAlaLeuLeuAlaLysLeuLysValSerGlySerPheIlePheTy 458
 Db 1475 CCTATTAGATGAGTGGTGGGAGTGGGATGATCATCTGTC-----ATGGCTTTCT 1525
 Qy 458 rLeuTyLeuPheValPheHisSerThrTrpValThrArgThrAlaTySerSerProSe 478
 Db 1526 TCTATCATCCTACAGTTTCATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1585
 Qy 478 rValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuLysAspPheArgG 498
 Db 1586 CATTTGTTCTGCTGCTGCTGGTGGGATGGCAGTGGATCATTTTGTGACTTCGAG 1645
 Qy 498 uAlaTyTrpTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAs 518
 Db 1645 ACCTATTATTGCTTCCCTCATATCTCTCAGAGATGCAAAAGTATGATGCTGTTGGGA 1705
 Qy 518 pTyGlyTyGlnIleGlyMetAlaAspArgThrThrLeuValAspAsnThrTr 538
 Db 1706 TTATGCTACCAAAATTACTGCAATGCAAAATCGCAATTTTATGTCACAAATACACATG 1765
 Qy 538 pAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTy 558
 Db 1766 GAATAATACCATATTTCTCGAGTAGGCGAGCATGGCATCCACAGCAAGAAAGCCTA 1825
 Qy 558 rGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleG 578
 Db 1826 TGAATATCATGAGGAGCTGTGATGTCAGTATGCTCTGTCTATTTTGGAGGCTTACTG 1885
 Qy 578 yPheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIle-- 597
 Db 1886 GTATTCTCGGATGATATCAACAGTTTCTTGTGATGCTCCGATGCTGGAGGAGCAGACA 1945
 Qy 598 -TrpProGluGluIleLysGluArgTyrPheTyThrAlaGluGlyGluTyArgValAs 617
 Db 1946 GACAGGAAGACACATTAAGGAGATGACTACTATCTCTCTCTCTCTCTCTCTCTCTCT 2005

Qy 617 pAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPh 637
 Db 2006 TCGTAGGGTTCCTCGGTGCTCTCACTGCTTGTATGTACAAATGTGTACTACCGCTT 2065
 Qy 637 eProGlnLeuPheAsnGlyGlyGln-----AlaThrAspArgValArgGlnGlnMe 654
 Db 2066 TGGGCAAGTCTACACAGAGCCAGCGTCCACAGGCTTTGACCGTGTTCGAAATGCTGA 2125
 Qy 654 tIleThrProLeuAspValProProLeuAspTyrPheAspGluValPheThrSerGluAs 674
 Db 2126 GATTGCTTAATAAGACTTT--GAGCTTGATGTCCTCGAGGACGATATACACAGACA 2182
 Qy 674 nTrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGly 689
 Db 2183 CTGCTAGTACAGATATACAGGTAAAGACCTGGATAATCGAGGC 2228

RESULT 6

BC028897 2656 bp mRNA linear HTC 04-AUG-2004
 LOCUS Mus musculus cDNA clone IMAGE:4168013, containing frame-shift errors.
 DEFINITION

ACCESSION BC028897.1 GI:22137711

VERSION BC028897

KEYWORDS HTC

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2656)

AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Scaletton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Chen, R.P., Frange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,

Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

Fahy, J., Helton, E., Kertman, M., Madan, A., Rodriguez, S.,

Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Souffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

Butterfield, Y.S., Krzyzinski, M.I., Skalska, U., Smal, D.E.,

Schneer, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2656)

Direct Submission

Director MGC Project.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

NTM-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: angbcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegod, H.,

Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,

A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAX Plate: 24 Row: F Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 31560679
This clone has the following problem: frame shifted.

FEATURES

Source

1. 2656

/organism="Mus musculus"

```

/motif type="mrna"

```

```
/usr/bin/nc -l 4444
```

```
/db xref="taxon:10090"
```

```
/usr_x86_64-linux-gnu/libc.so.6: error: /usr_x86_64-linux-gnu/libc.so.6: undefined symbol: __clone@GLIBC_2.1.2
```

/tissue type="Salivary gland, 10 week old female mouse"

```
/clone_type= salivary_gi
/clone_lib="NCI CGAP SG2"
```

```
/crome_tld=NOT_CGAF_592
/lab host="DH10B"
```

```

/ tab_host= DHT06
/ note="Vector: PCMV-SPORT6"

```

ORIGIN

Alignment Scores:		
Pred. NO.:	6.49e-204	2556
Score:	1925.50	378
Percent Similarity:	69.76%	107
Best Local Similarity:	54.39%	182
Query Match:	51.28%	28
DB:	3	10
	Length:	Matches:
		Conservative:
		Mismatches:
		Indels:
		Gaps:

US-10-028-384-6 (1-718) x BC028897 (1-2656)

QY	14	ThrLeuLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeu	33
DB	65	ACACTTCTAAAGCTTCTCATCTGTCGAGCTGCTGTATATCTTTTCTACTCGTCT	124
QY	34	PheAlaValIleLysPheGlySerIleIleHisGluPheAspProTyrPheAsnTyrArg	53
DB	125	TTTGCTGTGTCAGATTGAAAGTGCATCCATCAGTTTGATCGTACTTTAATATCG	184
QY	54	AlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnThrPheAspAspArg	73
DB	185	ACTACCGGTTTCTGGCTGAGAGGGGTTTATAAATTCATAACTGGTTGAAGACCG	244
QY	74	ThrTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrThr	93
DB	245	GCTTGGTACCCCTTGGCGCGAATCATTTGGAGGAACAATTTACCAGGCTTTAATGATCACT	304
QY	94	SerAlaPheIleThrPheAlaLeuArgAsnThrLeuGlyLeuProIleAspIleArgAsn	113
DB	305	TCTGTGCATCTACATGTACTC---CATTTCTCCATAACACTATGTACATTCGAAT	361
QY	114	ValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGluPhe	133
DB	362	GTCGTGTGTTTCTCGCGCCCATTTTTCCTCTTTCACCACCATCGTTAGTACCACTT	421
QY	134	ThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaGlyPheIleAlaIleVal	153
DB	422	ACCAAGAGCTCAAGATGCGAGGCTGGGCTTCTGCTGCGCATGATGCTGTAGTT	481
QY	154	ProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaThr	173
DB	482	CTGGGTATATTCTCGATCTGTAGTCGCTCCATATGTAATGAAGAAATGTATCTTT	541
QY	174	LeuLeuMetValThrPheMetPheTyrIleLysAlaGlnLysThrGlySerIleMetHis	193
DB	542	TGCATGCTGCTTACTACTACATGTGGATCAAGCAGTGAAGACTGTTCCATCTATTGG	601
QY	194	AlaThrCysAlaAlaLeuPheTyrThrMetValSerAlaTrpGlyGlyTyrValPhe	213
DB	602	GCTGCCAAGTGTGCCCTCGCTTATTCTACATGGTCTCTTCATGGGAGGCTATGTGTC	661
QY	214	IleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSerSer	233
DB	662	CTGATCAACTGATTCCTTACATGCTCGCTGGTGAATGTGTGACAGCGCGTTTCTCAC	721

QY 579 eGlycylAspSerIleuPheLeuTrpMetIleArgIleSerGluGlyle---Tr 598
 Db 1781 TTCTCGGATGATATCAACAGATTCTTTGATGTCCTCGGATTCGAGGACACAGAC 1840
 QY 598 pProGluGluIleLysGluArgTyrPheTyrThrAlaGluGluGlyLysTrpArgValAspAl 618
 Db 1841 AGGAGACACATTAAAGAGATGACTACTACTCTCTACTCTGCGGAATTCGGTGTTCATCG 1900
 QY 618 aArgAlaSerGluThrMetArgAnSerLeuLeuTyrLysMetSerTyrLysAspPhePr 638
 Db 1901 TGAGGGTTCTCGGTCGTCCTCAACTGCTTATGATACAAATGTTACTACCGCTTTCG 1960
 QY 638 cGlnLeuPheAsnGlyGlyGln-----AlaThrAspArgValArgGlnGlnMetIl 655
 Db 1961 GCGGTCTACACAGACGACGCTCCACAGGCTTTCACCGTGTTCGAATGCTGAGAT 2020
 QY 655 eThrProLeuAspValProProLeuAspTyrPheAspGluValPheThrSerGluAnTr 675
 Db 2021 TGGTAATAAGACTT---GAGCTTGATGTCCTCGGAGGAGCGTATACACAGAACTG 2077
 QY 675 pMetValArgIleTyrGlnLeuLysLysAspAlaGlnGly 689
 Db 2078 GCTACTCGGATATCAAGGTAAGGACCTGGAATTCGAGGC 2120

RESULT 7
 AK077877
 LOCUS
 DEFINITION
 Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
 enriched library, clone:6030401L06 product:OLIGOSACCHARYL
 TRANSFERASE STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
 homolog [Mus musculus], full insert sequence.

ACCESSION
 AK077877
 VERSION
 1
 KEYWORDS
 HTC; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5
 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 3643)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murota, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, K., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 Location/Qualifiers
 1. 3643
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM DB:6030401L06"
 /db_xref="taxon:10090"
 /clone="6030401L06"
 /sex="male"
 /tissue_type="testis"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="13 days embryo"
 93. -2209
 misc_feature
 /note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
 (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
 (SWISSPROT P46978, evidence: FASTA, 99.7% ID, 100% length,
 match=2115)
 putative"

ORIGIN
 Alignment Scores:
 Pred. No.: 1,096-202 Length: 3643
 Score: 1919.50 Matches: 385
 Percent Similarity: 68.84% Conservative: 112
 Best Local Similarity: 51.32% Mismatches: 191
 Query Match: 51.04% Indels: 35
 DB: 3 Gaps: 12
 US-10-028-384-6 (1-718) x AK077877 (1-3643)
 QY 14 ThrIleuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeu 33
 Db 138 ACATTCTAAAGCTTCTCATCTCGATCGCTGCTGTTATCTCTTTCTCTCTCTT 197
 QY 34 PheAlaValIleLysPheGluSerIleIleHisGluPheAspProTrpPheAnTrArg 53
 Db 198 TTTCCTGCTGAGATTGAAGTGTCATCCAGTTCGATCCGCTACTTTATATATCGG 257
 QY 54 AlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAnTrPheAspArg 73
 Db 258 ACTACCCGGTTCTCGGCTGAGGAGGGGTTTATAAATTCATTAATGATGATCGG 317

74 ThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrThr 93
Db GCTTGGTACCTTTGGGCGCAATCATTTGGAGGAAACAATTTACCCAGGTTTAAATGATCACT 377

94 SerAlaPheLeuTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsn 113
Db TGTGTCGCAATCATCACTGATCTC---CATTTCTTCATATCCACATGACATTCGGAAAT 434

114 ValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGluPhe 133
Db GTCGTGTGTTTCTCGCCCGCGCTTTCTCTCTCTTCCACCACCATCGTTAGCTACCACTT 494

134 ThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleVal 153
Db ACCAAGAGCTCAAGAGTCAAGAGCTGGGCTTCTGCTGCTGCATGATTCGTAGATT 554

154 ProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIleThr 173
Db CCTGGGTATATTTCTCGATCTGTAGCTGGCTCTTATGATATGATGAAGAAATTCCTATCTT 614

174 LeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMetHis 193
Db TCATGCTGCTTACTTACTATGATGAGATCAAGGAGTGAAGATGTTTCATCTATTGG 674

194 AlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrValPhe 213
Db GCTGCAAGTGTGCGCTCGCTTATTTCTACATGGTCTCTTCTCATGGGAGGCTATGTGTTT 734

214 IleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSerSer 233
Db CGGATCTACGTAGCTGCTACTGTTTACTGCTGCTGGGACCATTTCTTCTATGACGATT 854

254 ProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGlyValPhe 273
Db TCCTTTGTTGGTTTCCAGCGCGCTTTCATCAGAACACATGGCAGGCTTTGAGTGTTT 914

274 GlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLys 293
Db GGTCTCTGTGTCAGATCCATGCTTTCTAGATTACCTGGCGCAGCAAGTTGAATCCACAGCAA 974

294 PheLysValIleMet-----MetValSerLeu-----PheLeuIleLeuValLeuGly 309
Db TTCGAAGTCTTTCCGAGGTGTTATCTCCCTGGTGGCTTTGCTTCTTCTCTCCTCCTCCT 1034

310 ValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPhe 329
Db GCTCTC-----CTCATGCTAAACAGGAAAAATTTCTCCCTGGACAGGCGGCTTTC 1082

330 TyrSerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSer 349
Db TACTCTCTGCTGATCCCTCTTATGCTAAGATTAACATTCCTCATATGATGATCTGTTTCT 1142

350 GluHisGlnProValSerTrpProAlaPhePhePheAspThrHisPheLeuIleTrpLeu 369
Db GAGCACCAGCCACCAACCTGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1202

370 PheProAlaGlyValPheLeuPheLeuAspLeuLysAspGluHisValPheValIle 389
Db TTTCCAGTGGCTCTTATCTGCTTTAGCAACCTGCTGCTGATGCTCGGATTTTATCATC 1262

390 AlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 409
Db ATGTATGGTGTGACCAAGATGATCTTTTCTGCTGTTTCTGCTGCTGCTGCTGCTGCTGCT 1322

410 ThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyr--- 428
Db GCACCTGTTATGTCATCTTCTTCTGCGATGTTGTTTCTCCAGGTGCTGCTCAGATATATG 1382

429 -----LeuAspPheLysThrSerAspArgLys-----Tyr 438
Db AAAAATCTGGACATAAGTCCGCCAGACACAGAGCAAGCAAGCAAGGATTCTACTTAC 1442

439 AlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyr 458
Db CCTATTAGATGAGTGGCGAGTGGATGATACTGGTC-----ATGGCTTCTTT 1493

459 LeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerProSer 478
Db CTCTATCACTACAGTCTTCTTCTGACTGGGTCAGCCAGTGAAGCCCTATTCTTCTCCCTCC 1553

479 ValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGlu 498
Db ATTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1613

499 AlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAsp 518
Db CGGTATTATTGGCTCCGTCACAACTACTTCCAGAGGATGCAAAAGTCATGTCATGGTGGAT 1673

519 TyrGlyTyrGlnIleGlyMetAlaAspArgThrThrLeuValAlaAspAsnAsnThrTrp 538
Db TATGGCTACCAAAATTTCTGCAATGGCAATCGGAA-ATTTAGTGAGCAATTAACATGG 1732

539 AsnAsnThrHisIleAlaIleValGlyLysAlaValMetAlaSerProGluLysSerTyr 558
Db AATAATACCCATATTTCTCGAGTAGGCGAGCAATGGCATCCACAGAAAGAAAGCCTAT 1792

559 GluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGly 578
Db GAAATCATGAGGAGCTTGT 1852

579 PheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIle--- 597
Db TATTCTTCGGATGATATCAACAAGATTTCTTTGGATGCTCCGGATTCGAGAGCAACAGAG 1912

598 TrpProGluGluIleLysGluArgTyrThrAlaGluGlyGluTyrArgValAsp 617
Db ACAGAAAGACACATTAGGAGAGATGACTACTACTACTACTACTACTACTACTACTACTACT 1972

618 AlaArgAlaSerGluThrMetArgAsnSerLeuTyrLysMetSerTyrLysAspPhe 637
Db CGTGAGGCTTCTCCGGTGTGCTCACTGCTTATGTACAAAATGTGTACTACCGCTTT 2032

638 ProGlnLeuPheAsnGlyGlyGln-----AlaThrAspArgValArgGlnGlnMet 654
Db GGGCAGGTCTACACAGAGCCAGCGCTCCACAGGCTTTGACCGTCTTCGAATGCTGAG 2092

655 IleThrProLeuAspValProProLeuAspTyrPheAspGluValPheThrSerGluAsn 674
Db ATTGGTAAATAAGACTTT---GAGCTTGTATGCTCTGGAGAGCGGTATATACCAAGAACAC 2149

675 TrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGly-----ArgThr-Le 692
Db TGGCTAGTCAAGATATACAGGTAAAGAGCTGATTAATCGAGCTTGTCAAGACATAA 2209

692 uArgAspValGlyGluLeuThr-----ArgSerSerThrLysThrArgArgSerIle 709
Db ACCTCACATTGGTCCCTGAGCATTTATGCTTCGACTGAGCGGCTCATGTTGAGGACGCT 2269

709 eLys 710
Db GAAG 2273

RESULT 8
AY418284
LOCUS
DEFINITION Homo sapiens ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
Genomic survey sequence.
ACCESSION AY418284
VERSION AY418284.1 GI:39774244
KEYWORDS GSS.
SOURCE Homo sapiens (human)

QY 34 PheAlaValIleLysPheGluSerIleHisGluPheAspProTrpPheAsnTyrArg 53
 DB 213 TTTGCTGCTGAGATTGAAGTGTCTCATCTGAGTTGATCGTACCTTAAATTATCGG 272
 QY 54 AlaThrIysTyrLeuValAsnAsnSerPheTyrIysPheLeuAsnTrpPheAspArg 73
 DB 273 ACTACCGGTTTCTGGCTGAGGAGGGGTTTATAAATTCATTAACCTGGTTGATACCGG 332
 QY 74 ThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrThr 93
 DB 333 GCTTGGTACCTTTGGGCGGATGATTGGAGGACAAATTTACCGAGTTTATGATCACT 392
 QY 94 SerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsn 113
 DB 393 TCTGTGCAATCTACCATGTACTC---CATTTCTCCATATCATCACTACCATTCGGAAT 449
 QY 114 ValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGluPhe 133
 DB 450 GTCTGTGTTTCTGGGCGGCTTTCTCTCTTTTACCACCATCTGTTAGTACCACTT 509
 QY 134 ThrIysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaVal 153
 DB 510 ACCAAAGAGCTCAAGGATGCAAGGAGCTGGGGTTCTTGGTGGTGCATGATTCTGTAGTT 569
 QY 154 ProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaThr 173
 DB 570 CTTGGGATATTTCTCGATCTGAGCTGGCTCTTATGATTAAGAGGATTCGTATCTTT 629
 QY 174 LeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMetHis 193
 DB 630 TGCATGCTGCTTACTTACTCATGTGATCAAGGAGCTGAAGAGCTGGTTCCATCTATTGG 689
 QY 194 AlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrValPhe 213
 DB 690 GCTGCCAAGTGGCCCTCGCTTATTTCTACATGGTCTCTTCATGGGAGGCTATGTTTC 749
 QY 214 IleThrAsnLeuIleProLeuHisValPheLeuLeuLeuLeuMetGlyArgTyrSer 233
 DB 750 CTGATCACTGATTCCTCTACATGCTCTGCTCTAATGCTGACAGCGGTTTCTCAC 809
 QY 234 LysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIle 253
 DB 810 CGGATCACTAGCTAGCTACTCTGTTTACTGCTGGCACCACTTCTTCTATGAGATT 869
 QY 254 ProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaLeuGlyValPhe 273
 DB 870 TCCCTTGTGTGTTTCCAGCGCGTCTTCTATCAGAACACATGCCAGCCCTTGGAGTGT 929
 QY 274 GlyLeuIleGlnIleValAlaPheGlyAspPheValIysGlyGlnIleSerThrAla-Ly 293
 DB 930 GGTCTCTCATCATCCATGCTTTCTGATGATTACTGGGCACACAGTTGAATCCACACA 989
 QY 293 sPheIysValIleMet-----MetValSerLeuPheLeuIleValLeuGlyValVa 311
 DB 990 ATTCCGAAGTTTTCGCGAGTGTATCTCCCTGGTGTGCTTCTCCCGCCACACTGTGGG 1049
 QY 311 lGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyr-Se 331
 DB 1050 ACCCTCCCATGCTACACA-----GGAAAAATTTCTCCCTGGACAGCGGCTTTTACTC 1103
 QY 331 rLeuTrpAspThrAsnTyrAlaIysIleHisIleProIleIleAlaSerValSerGluHi 351
 DB 1104 TCTGCTGAATCCCTCTAATGCTAAAAAACAATCCCATTTATGGCATCTGTTCTTGAGCA 1163
 QY 351 sGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPhePr 371
 DB 1164 CCAGCCCAACAACCTGGCTTCTTACTATTATTGATCTACAGCTCTCTGCTTCTCATGTTCC 1223
 QY 371 oAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyr 391
 DB 1224 AGTTGCCCTCTATTACTGCTTTTACCAACCTGCTGATGCTCGGATTTTATCATCATGTA 1283
 QY 391 rSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrPr 411

DB 1284 TGGTGTGACCAAGCATGTACTTTTTCAGCTGTAATGGTGGTCTAATGCTGGTATTGGCACC 1343
 QY 411 oValIleCysValSerAla-AlaValAlaLeuSer-----L 423
 DB 1344 TGTATTGTGCATCTCTTTCTGGGCATTTGGTGTTCCTCCAGGGTGGCTGTCACATATANGA 1403
 QY 423 ysIlePheAspIleTyrLeuAspPheLysThrSer-----AspArgLysTyr 438
 DB 1404 AAAATCTGGACATAAGTCGCCACACAAGAGAGCAAGAACCAAGGATCTTACTTAC 1463
 QY 439 AlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyr 458
 DB 1464 CCTATTAAAGAAATGAGGTGGCGAGTGGGATGATCTGCTC-----ATGGCTTCTCTT 1514
 QY 459 LeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAla-TyrSerSerProSe 478
 DB 1515 CTATCACCTACACGTTTTCATTCGACTTGGGTGACAGTGAAGCCCTATCTTCTCCCTC 1574
 QY 478 rValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuAspPheArgG 498
 DB 1575 CATTTGCTGCTGCTGCTGGTGGGATGGCAGTAGGATCATTTTGTGATGATTCGAGA 1634
 QY 498 uAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAs 518
 DB 1635 AGGTATTATTGGCTCCGTCACATACTCCAGAGGATGCAAAAGTCTGATGTGGGA 1694
 QY 518 pTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnThrThr 538
 DB 1695 TTATGCTCTACCAAAATTTACTGCAATGGCAATCGGACAAATTTTAGTGACAAATACACATG 1754
 QY 538 pAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGlu-GluLysSerT 558
 DB 1755 GAATATATCCCATATTTCTCGAGTAGGCGAGCAATGSCATCCACAGAAAGAAAGCCT 1814
 QY 558 rGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleG 578
 DB 1815 ATGAATCATGAGGAGGCTGATGTCAGTATGCTGTGTCATTTTGGAGGCTTACTG 1874
 QY 578 lPheGlyGlyAspAspIleAsnLysPheLeuTyrMetIleArgIleSerGluGlyIle- 597
 DB 1875 GGTATTCTTCGGATGATATATCAACAGTTTCTTGGATGGTCCGATTTGGAGGAGCACAG 1934
 QY 598 --TrpProGluIleLysGluArgTyrPheTyrThrAlaGluGlyLysTyrArgVala 617
 DB 1935 AGACAGGAAGACACATTAAGGAGAAATGACTACTACTACTCTGCTGCGGAATTCGGTGTG 1994
 QY 617 sAlaAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAsp 637
 DB 1995 ATCGTGAGGGTTCTCCGGTGTGCTCAACTGCCITATGTACAAATGTGTACTACCGCT 2054
 QY 637 heProGlnLeuPheAsnGlyGlyGln-----AlaThrAspArgValArgGlnLm 654
 DB 2055 TTGGGAGGCTCTACACAGAACGAGCGTCCACAGGCTTTGACCGGTTTTCGAAATGCTG 2114
 QY 654 eTyrThrProLeuAspValProProLeuAspTyrPheAspGluValPheThrSerGluA 674
 DB 2115 AGATTGGTAAATAAGACTTT--GAGCTTGTATGCTCTGGAGGAGGCTATATACACAGAAC 2171
 QY 674 snTrpMetValArgIleTyrGlnLeuLysAspAspAlaGlnGly 689
 DB 2172 ACTGGCTAGTCAGGATATACAGGTAAGGCTGGATATATCGAGGG 2218

RESULT 10

AY418286 2103 bp DNA linear GSS 17-DEC-2003
 LOCUS Mus musculus ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
 DEFINITION genomic survey sequence.
 ACCESSION AY418286
 VERSION AY418286.1 GI:39774246
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus


```

|||||.....|
1677 GAAGAAAAGCTATGAATCATGAGGAGCTTGATGTGAGCTATGCTGCTGATTTT 1736
QY GlyGlyLeuLeuGlyPheGlyGlyAspAlaAsnLysPheLeuTrpMetIleArgIle 593
Db .....|
1737 GGAGGCTTACTGGGTATCTCGATGATATCAACAGTTTCTTGGATGCTCGGATT 1796
QY SerGluGlyIle---TrpProGluGluLeuGlyGlyArgGlyTrpPheTrpThrAlaGluGly 612
Db .....|
1797 GGAGGAACACAGAGACAGAGACACATTAAGAGAGATGACACTATATCTCTACTCGG 1856
QY GluTrpArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTrpLysMet 632
Db .....|
1857 GAATTCGGTGTGATCGTGAGGGTTCCTCGGTCTCTCACTGCTTATGTACAAATG 1916
QY SerTrpLysAspPheProGlnLeuPhe---AsnGlyGly-GlnAlaThrAspArgValArg 651
Db .....|
1917 TGTACTACCCCTTGGCAGGCTCATACAGAGCGCTCCACAGCGCTTTGACCGTTCG 1976
QY gclnlnMetIleThrProLeuAspValProProLeuAspTrpPheAspGluValPheTh 671
Db .....|
1977 AATGCTGAGATTGGTAATAAGACTTT---GAGCTTGATGCTCTCGAGGAGCGTATAC 2033
QY rSerGluAsnTrpMetValArgIleTrpGlnLeuLysLysAspAspAlaGlnGly 689
Db .....|
2034 CACAGAACCTGGCTAGTCAGGATATACAAAGTAAGGACCTGGATATCGAGGC 2088

RESULT 11
CNSOADYZ 2419 bp mRNA linear HTC 06-FEB-2004
LOCUS Arabidopsis thaliana Full-length cDNA Complete sequence from clone
DEFINITION GSITPGH512C09 of Hormone Treated Callus of strain col-0 of
Arabidopsis thaliana (thale cress).
ACCESSION BX816490
VERSION BX816490.1 GI:42474670
KEYWORDS HTC; GSLT cDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 2419)
Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C.,
Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome
Annotation
Unpublished
2 (bases 1 to 2419)
Genoscope.
Direct Submission
Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out
full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli
V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C.,
Schachter V., Weissenbach J., Salanoubat M.
URGV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis
genome released by MIPS (Munich Information center for Protein
Sequences). 5 prime and 3 prime are assembled with Phrap.
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
length
http://www.genoscope.cns.fr/cgi-bin/ggb?source=Arabidopsis.
Location/Qualifiers
1..2419
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Col-0"
FEATURES
source

```

```

/db_xref="taxon:3702"
/clone="GSLTPGH512C09"
/tissue_type="Hormone Treated Callus"
/plasmid="pCMVSPORT_6"
1..2419
/gene="Atlg34130"
ORIGIN
Alignment Scores:
Pred. No.: 3,67e-177 Length: 2419
Score: 1689.00 Matches: 337
Percent Similarity: 65.14% Conservative: 119
Best Local Similarity: 48.14% Mismatches: 206
Query Match: 44.91% Indels: 38
DB: 3 Gaps: 11

US-10-028-384-6 (1-718) x CNSOADYZ (1-2419)
QY 13 GlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArg 32
Db .....|
222 GAGCTCTTACTCGTATCTCGATCTCTGCTCTCATCTCTAGCTTTTCATAGCTCGT 281
QY LeuPheAlaValIleLysPheGluSerIleHisGluPheAspProTrpPheAsnTrp 52
Db .....|
282 CTCTTCAGTGTCTCGCTACGATCAATGATCAAGATTCGATTCGATTCGATTCACATC 341
QY 53 ArgAlaThrLysTrpLysLeuValAsnAsnSerPheTrpLysPheLeuAsnTrpPheAspAsp 72
Db .....|
342 CGTACGACCTTTCTCTCTCACAGAGAAAGCTTTTACGAGTTTGGAACTGGTTCGATTC 401
QY 73 ArgThrTrpTrpProLeuGlyArgValThrGlyGlyThrLeuTrpProGlyLeuMetThr 92
Db .....|
402 GAGAGTTGATACCTTAGTTCGATCATCGCGGTACTCTTTATCCGGTCTTATCGGTC 461
QY 93 ThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProLysIleAspIleArg 112
Db .....|
462 ACAGCTGCGCTTAATCTACTGGACATTCGGA---TTTCTCCGATTTCTTCGTTTCATATCCGT 518
QY 113 AsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTrpGlu 132
Db .....|
519 GAGTTTGTGTGTTAACCGCACCCTTTTCGCGTCGATACGACTCTCGTTCGATTTTC 578
QY 133 PheThrLysGluLeuLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIle 152
Db .....|
579 TTTGGTAAAGAGCTTTGGGATACAGAGCTGCTCTTGTGTCTGTCTTCTTATCGCTATT 638
QY 153 ValProGlyTrpLysSerArgSerValAlaGlySerTrpAspAsnGluAlaIleAlaIle 172
Db .....|
639 TGTCCGGATACATTTCTGTTCTAGTCTGGTTCGATGATATACGAGCTGTTCGATT 698
QY 173 ThrLeuLeuMetValThrPheMetPheTrpLysAlaGlnLysThrGlySerIleMet 192
Db .....|
699 TTGCTCTCTCTGCTGACCTTTTATCTTTGTTAAAGCGGTGAACACTGGTTCATTAGCT 758
QY 193 HisAlaThrCysAlaAlaLeuPheTrpPheTrpMetValSerAlaTrpGlyGlyTrpVal 212
Db .....|
759 TGGCTCTGCTCTGCTGCTTGTGTTTATCTTTATGTTTCAGCTTCGGGAGGATATG 818
QY 213 PheIleThrAsnLeuLysProLeuHisValPheLeuLeuLeuLeuMetGlyArgTrpSer 232
Db .....|
819 TTATTTATCAATCTGCTCTCTCTTTACGTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 878
QY 233 SerLysLeuTrpSerAlaTrpThrTrpTrpValIleGlyThrValAlaSerMetGln 252
Db .....|
879 ATGAGCTTTTACATGCTTATTAACGTATGATATCTTGGGGATGTTGCTCGCATCAA 938
QY 253 IleProPheValGlyPheLeuProLeuArgSerAsnAspHisMetAlaAlaLeuGlyVal 272
Db .....|
939 ATTGCTTCTGCTGTTTC---ACATGTTTCATCTGTTGATATGTTGTTGTTGTTGTTGTTGTT 995
QY 273 PheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAla 292
Db .....|
996 TTCTTATTATGACAGGTGTTTACTTCTTGGACTGGGTTTAAAGTACCAGCTCAATGACACC 1055

```


293 Lys--PheLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyVal 311
 1056 AAGTTGTCACAAACGTTTTGAGATAAAGTGGACATCGGCTATTCGTGGTGGT 1115
 312 GlyLeuSerAlaLeuThrThrMetGlyLeuIleAlaProThrGlyArgPheTyrSer 331
 1116 GCTGGGGTGGTGGACAGCTCTGGCTATATATCCATGAGCTGGTGGATTTACTCG 1175
 332 LeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHis 351
 1176 TTGCTTGACCCGCACTTACGCAAGGATCATATTCCTATTATGATCTGTTCTGAGCAT 1235
 352 GlnProValSerTrpProAlaPhePhePheSerThrHisPheLeuIleTrpLeuPhePro 371
 1236 CAGCTTACCTGCTGGTGGTATCATGTTGACATATATTCCTTCTTCTTCTTCTTCT 1295
 372 AlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyr 391
 1296 GCGGGTCTTACTTCTGTTTCAAGCGTTTACGGATGCTAGCATATTTATTTGTCATGAT 1355
 392 SerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrPro 411
 1356 GGTCTCAGTACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
 412 ValIleCysValSerAlaAlaValAlaLeuSer-----LysIlePheAspIleTyr 428
 1416 GCAGTTTGTCTTATCAGTGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1475
 429 LeuAspPheLys----- 432
 1476 TTAAGCAAAACAAAGGTTTCTCAGACTGTTCCAGCAAGAGCTGGTAGTTCACAA 1535
 433 -----ThrSerAspArgLysTyrAlaIleLysProAlaAlaLeuLeuAla 1547
 1536 GCTTCTTCAAGAGTTACACTTTCATGCTCAGCTCAGCTTCCAGAGAATGGTGCCATGCT 1595
 448 LysLeuIleValSerGlySerPheIlePheTyrLeuTyr---LeuPheValPheHisSer 466
 1596 CTTCTTGTGTT-----GGTGTATTTATTTGCTCAGTATGATGCTATTCATGCT 1643
 467 ThrTrpValThrArgThrAlaTyrSerSerProSerValValLeuProSerGlyThrPro 486
 1644 ACATGGCTCAGCAGCAGGATATTCATCTCCCTCAATGCTTAGCTGCAAGAGGAGCC 1703
 487 AspGlyLysLeuAlaLeuIleAspPheArgGlyAlaTyrTyrTrpLeuArgMetAsn 506
 1704 CACGGGAACAGAAATCATCTTTGATGATACCGGAGGCACTACTGCTTAGGCAAAAC 1763
 507 SerAspGluAspSerLysValAlaAlaTrpTrpAspTyrGlyTyrGlnIleGlyMet 526
 1764 ACTGCCACTGATGCTAAGATTATGCTGGTGGGACTATGGTACCAATCCTGCTCCATG 1823
 527 AlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleVal 546
 1824 GGAATAGAACTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1883
 547 GlyLysAlaMetAlaSerProGluGluLysSerTyrGluIleLeuLysGluHisAspVal 566
 1884 GGACGCCCATGCTTCTTATCAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1943
 567 AspTyrValLeuValIlePheGlyLeuIleGlyPheGlyGlyAspIleAsnLys 586
 1944 AACTATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2003
 587 PheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluIleLysGluArgTyr 606
 2004 TTTCTTGGATGGTGGAGATTGGAGTGGAGTATTTCCCA---GTGATCAAGACCT--- 2057
 607 PheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAlaSerGluThrMetArgAsn 626
 2058 GATTACCTTGTGAATGGTGAATTCCTGCTAGACAAAGGTGATCACCAGAGATGCTGAAC 2117

627 SerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu-----PheAsnGly 643
 2118 TGCTCTCATGTACAAAGTTATGCTATTACAGGTTCGGTGGCTGACCAAGATACGCAAG 2177
 644 GlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspValProLeu 663
 2178 CCACCCGGGTATGATAGCAAGAGGAGTGGAGTACGGACCAAGATATC---AAACTT 2234
 664 AspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyrGlnLeuLys 683
 2235 GAACACTTAGAAGAGCTTACACAACTCCCACTGAGTATGCTGCTTATTTACAGAGTAA 2294
 RESULT 12
 AY418285 2079 bp DNA linear GSS 17-DEC-2003
 LOCUS Pan troglodytes ITMI gene, VIRTUAL TRANSCRIPT, partial sequence,
 DEFINITION genomic survey sequence.
 ACCESSION AY418285
 VERSION AY418285.1 GI:39774245
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 REFERENCE 1 (bases 1 to 2079)
 AUTHORS Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 2079)
 AUTHORS Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Direct Submission
 JOURNAL Submitted (18-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 Location/Qualifiers
 1..2079
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 <1..>2079
 /gene="ITMI"
 /locus_tag="HCM6506"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,49e-152 Length: 2079
 Score: 1465.00 Matches: 304
 Percent Similarity: 57.43% Conservative: 71
 Best Local Similarity: 46.55% Mismatches: 246
 Query Match: 38.95% Indels: 32
 DB: 9 Gaps: 7
 US-10-028-384-6 (1-718) x AY418285 (1-2079)
 50 PheAsnTyrArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnTrp 69
 154 TTTAATTATCGACTACCAAGNNCTGGCTGNNNGGGTTTATAAATCCATACTGG 213
 70 PheAspAspArgThrTrpTyrProLeuGlyArgValThrGlyLeuTyrProGly 89
 214 TTTGATGACCGAGCTGTGCTTGGAGATCATTTGAGGAGNNNTTACCAGGT 273
 90 LeuMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIle 109

Db 274 TTAATGATCACTCTGCTGCAATCTACCATGTACTC---CATTTTTTCCACATCCACCATC 330
 Qy 110 AspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAla 129
 Db 331 GACATTCGGAATGCTGTTGTTGCTCTGCGCCCTCTCTTCTCTCTTCCACCATCTGTC 390
 Qy 130 ThrTyrGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaGlyPhe 149
 Db 331 ACGTACCACTTACCAAGAGCTCAAGGATGCAAGGCTGCGGCTCTCTGCTGCTGCTG 450
 Qy 150 IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAla 169
 Db 451 ATTGCTGTAGTCTCTGATATATCTCCGATCTGTGCTGCTCTCTATATATATAGAGGG 510
 Qy 170 IleAlaIleThrLeuLeuMetValThrPheVerPheThrIleLysAlaGlnLysThrGly 189
 Db 511 ATTGCCATCTTTTGCATGCTACTCACCTACTACATGTGGATCAAGCAGTAAAGACTGCT 570
 Qy 190 SerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGly 209
 Db 571 TCCATCTCTGGGAGCTAGTGTGCTCTGCTTATTTCTACATGTCTCGTCATGGGA 630
 Qy 210 GlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuMetGly 229
 Db 631 GCTTATGTCTCTGATCACTCACTTAATCTCTCCAGCTCTCTGCTGCTGCTGCTCAAGGC 690
 Qy 230 ArgTyrSerSerLysLeuTyrSerAlaTyrThrTyrTrpTyrAlaIleGlyThrValAla 249
 Db 691 GCTTCTCTCAGGAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 750
 Qy 250 SerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAla 269
 Db 751 TCTATGAGATCTCTCTTGTGGGTTTCCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 810
 Qy 270 LeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIle 289
 Db 811 NNN 870
 Qy 290 SerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuIleValLeuGly 309
 Db 871 AATCCANAACAATTNNNGTTCTTTCCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 930
 Qy 310 ValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPhe 329
 Db 931 NNN 990
 Qy 330 TyrSerLeuTrpAspThrAsnTyrAlaLysIleIleIleProIleLeuAlaSerValSer 349
 Db 991 TACTCACTGCTGGATCTCTTCTTATGCTTAAGAACACATCCCATCATCTCTGCTGCT 1050
 Qy 350 GluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeu 369
 Db 1051 GAGCATAGCCCAACACTGCTCTCATATATTTTGGCTNNAGCTCTGCTGCTCAAG 1110
 Qy 370 PheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIle 389
 Db 1111 TTTCAGTGTGCTCTTACTGCTTNGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1170
 Qy 390 AlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 409
 Db 1171 ATGTATGTGTGACCATGCTACTTTTCAGCTGTGTAATGNNNNNNNNNNNNNNNNNNNN 1230
 Qy 410 ThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu 429
 Db 1231 NNN 1290
 Qy 430 AspPhe-----LysThrSerAspArgLysTyr 438
 Db 1291 AAGAAATGACATAAANN 1350
 Qy 439 AlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValSerGlyPheThrPheTyr 458
 Db 1351 CCTATTAAAGATGAAGTGGGCAAGTGGGATGATATGCTC-----ATGCTTCTTCTT 1401

Qy 459 LeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSer 478
 Db 1402 CTCATCACCTACACCTTTTCATTCACCTGGGTGAGCAGTGGGCTACTCTTCTCCGCTCC 1461
 Qy 479 valValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGlu 498
 Db 1462 ATTGTACTATCTCCCCGGGTGGGATGCGAGTAGGATCATATTTTGTATGACTTCCCGAGA 1521
 Qy 499 AlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTyrTrpAsp 518
 Db 1522 GCATATATTTGCTTCGTCACATACTCCAGAGGTGCGAGGTCAATNNCTGTGGGAT 1581
 Qy 519 TyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrp 538
 Db 1582 TATGGGTATCATAGTACAGCTATGGCAACCCGACAAATTTTAGTGGCAATAACACATGG 1641
 Qy 539 AsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyr 558
 Db 1642 AATAATACCATATTTCTNN 1701
 Qy 559 GluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGly 578
 Db 1702 GAGATCATGAGGAGCTCGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1761
 Qy 579 PheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIle---SerGluGlyIle 597
 Db 1762 TATTCCTCTGATGATATCAACAAGTTCTNNNGATGTCGCGATGGAGGGAGNACAGAN 1821
 Qy 598 TrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAsp 617
 Db 1822 NNANGCAACATATCAAGGAGATGACTATTATATCTCAACTGGGAGTTCCTGCTGCTG 1881
 Qy 618 AlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPhe 637
 Db 1882 CNTGAAGTTCCTCAGCTGCTGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1941
 Qy 638 -----ProGlnLeuPheAsnGlyGlyGlnAlaThrAsp 648
 Db 1942 NNN 1983
 Qy 649 ArgValArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyrPheAspGlu 668
 Db 1984 CGTGTCCGAATGCTGAGATTGGGATAAAGACTTT---GAGCTTGATGTCTCGGAGAA 2040
 Qy 669 valPheThrSerGluAsnTrpMetValArgIleTyrGln 681
 Db 2041 GCATATACCAAGAACATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2079

RESULT 13

CNS06CLU/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

1088 bp DNA linear GSS 14-JUN-2001
 T7 end of clone AROAA006C05 of library AROAA from strain CBS 732 of
 Zygosaccharomyces rouxii, genomic survey sequence.

AL392680
 AL392680.1 GI:12142260
 GSS.

Zygosaccharomyces rouxii
 Zygosaccharomyces rouxii
 Zygosaccharomyces rouxii
 Zygosaccharomyces; Saccharomycetaceae; Zygosaccharomycetes;

1 (bases 1 to 1088)
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
 Malbert,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Sautin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)

20584711
 11152876


```

646 -----TACAGGGTCTCTGGCGGGAGGTTATGTGTTTCATCATCAACCTCATCCCT 696
220 LeuHisValPheLeuLeuLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyr 239
697 CTCCTATGTTTGTGCTGCTCTCTGATGAGAGGACAGAGAGAGAGAGAGAGAGAGAGAGAG 747
240 ThrThrTyrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 259
747 -----
260 ProIleArgSerAsnAspHisMetAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaVal 279
748 -----ATAGGTGCTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777
280 AlaPheGlyAspPheValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 299
778 GCGTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
300 ValSerLeuPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 317
838 -----TTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891
318 TyrMetGlyLeuLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 337
892 TAC -----ACAGG-TAT 902
338 AlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpPro 357
903 GCGAAATAATACATCCAAATATTGTCATCAGTGTCTGAACATCAGCTCAGCATCAGCATCAGCAT 962
358 AlaPhePheAspThrHisPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 377
963 TCTTTCTCTTGTGATCATCATATTCTTGTGTGATCATCTCCAGCAGGCTGATGTTGTTGCTGCTG 1022
378 PheLeuAspLeuLeuAspGluHisValPheValIleAlaTyrSerValLeuLeuLeuLeuLeu 397
1023 ATCAAAAATATCAAGATGAAAGATGATTT ----- 1052
398 PheAlaGly-ValMetValArgLeuMetLeuThr -----le 409
1053 -----GGCAGGGAAGTGCAGAAAGCATGTGACGGAGCAGGAGAGAGAGAGAGAGAGAGAGAG 1106
409 uThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrIle 429
1107 GGCCCCCAACATC ----- 1119
429 uAspPheLysThrSerAspArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLys 449
1120 -----AAGAG 1124
449 uIleValSerGlySerPheIlePheTyrLeuTyrLeuPheValPheHisSerThrTyrVa 469
1125 CATCGTGCACCATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCTCATGCT 1184
469 lThrArgThrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyLys 489
1185 CACCAACAGCCCTACTCCAGCCCCAGTGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1244
489 sLeuAlaLeuLeuAspAspPheArgGluAlaTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 509
1245 CAGAAATATCTTAGATGATTTAGAGAGCATCTTTGTTGAGAGCAAAACACGAGTGA 1304
509 uAspSerLysValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 529
1305 ACAGCCAGGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1364
529 gThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAl 549
1365 GACTACTCTGATAGCAATAACACCTGGAACACACACCATTCGCTGCTGCTGCTGCTGCTGCTGCT 1424
549 aMetAlaSerProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrVa 569
1425 TATGCTCTTCTAATGAAACACCCGCTTATAAATCATGAGGTCTCTTGTGTTGATTATGT 1484

```

```

/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
232..2334
CDS
/note="liver regeneration related protein LRRG109"
/codon_start=1
/product="Ac1573"
/protein_id="AAP78765.1"
/db_xref="GI:32264637"
/translation="MTAGLHMLNTLITVHSDVFLAPFTSGLTSLTFLTR
ELWQAGLACIAIPGVYISRSVAGSDEGIAIFALQFTYILMHSVGLCPRA
HRSPFGSSGSLTISLVAALPSGALVHVARPELYRVSAGVYPIINLIPLHVFV
LLLMRYKRYVIGFALLQAYFLYDLRLTKQERFTLFFGVSLAAGAFSLVYI
LTYTQMRYTQLHQCLNISLRHGLSSLIYFLCVPQQOAGSASKISTMKEYLAG
KVRHTVQEPKEEGELGNIKISIVTLMMLMFAVCTWVTSNVSSPVSLASYN
HDTNRIIDDFREAYFMRQNTDEHARVMSWDYQIAGMANRTLLVDNNTWNHSHI
ALVGKMSNSTEAKVRSIDVDVLIYFGVIGYSGDDINDKFLMVRVLAEGHPKD
IRGCDYFTQCEFRVDRAGSPTLNCIMYKMSYRFGEMOLDPRTPPGFDRTNAGIG
NKDIKFLAEAFTEHNLVRIYKYKADNRETIGHKPRVTNIPPKOKYLSKKDLAS
TEERPGHTQGLTVUSEEQKGRWLSVSGILGSEENVLPDCCQHSHSDHNLHQEQQ
NMTDLGLGSQVSAKGSFPQDRAPFSDRDFNRLCPFLSHWDSMNQNGNVRKMT
ESREQGAQDPGQPRV"
ORIGIN
Alignment Scores:
Pred. No.: 2,48e-145 Length: 2334
Score: 1404.50 Matches: 301
Percent Similarity: 57.42% Conservative: 90
Best Local Similarity: 44.20% Mismatches: 144
Query Match: 37.34% Indels: 147
DB: 3 Gaps: 12
US-10-028-384-6 (1-718) x AY310157 (1-2334)
QY 50 PheAsnTyrArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnTrp 69
DB 109 TTTAACTATAGTCACTCACTCATCTTGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 168
QY 70 PheAspAspArgThrTrpTyrProLeuGlyArgValThrGlyGlyLeuTyrProGly 89
DB 169 TTGATGAAGAGCATGTTGACCATCAGGAGAGATGCTGGTGGGACCGCTTATCCAGGA 228
QY 90 LeuMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIle 109
DB 229 TTGATGATACAGCTGCGCTTATT--CATTTGATTTTAAATACATTGAACATACAGCTT 285
QY 110 AspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAla 129
DB 286 CACATCAGAGATGATGTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 345
QY 130 ThrTyrGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPhe 149
DB 346 ACCTTCTCTACTAGAGACTTTGGAAACCGAGGAGGAGGCTCTTGTGCTGCTGCTGCTGCTGCT 405
QY 150 IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAla 169
DB 406 ATTGCTATTGTATCCCGGGTACATATCTCGTCCGTAGCGGGATCCTTTGATATACAGAGGC 465
QY 170 IleAlaIleThrLeuLeuMetValThrPheMetPheTrp ----- 182
DB 466 ATTGCGATTTCGCATCTTCAGTTCATCTACTTATGCGGCCATTCGCGGCTGCTGCTGCTGCT 525
QY 182 ----- 182
DB 526 TGCTTTCTGCCCATCGCTCTCCAGGCCCTTTGGTTCCTCAGGGCAITCACTCCACCATC 585
QY 183 -----IleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeu 199
DB 586 TCCTTGTGCGCGCTGCTCTTCTAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 645
QY 200 PheTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuLeuPro 219

```

QY 569 lleuValillepheGlyLeulleGlypheGlyAspAspileAsnlySpheLeuYr 589
 Db 1485 GTTGTCTATTGAGGAGTATTGGTCTTCTGGTATGATATCAACAGTTCCTCTG 1544
 QY 589 pMetileArgileSerGluGlyIleTrrProGluGlyLeulleGlyArgTrrPheYrTh 609
 Db 1545 GATGTCAGATAGCTGAGGGGACACCCAAAGACATTCGGAGGCGNACTTTCAC 1604
 QY 609 rlaGluGlyGluTrrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeule 629
 Db 1605 CCAGCAGGAGAGTTCGAGGTGGACAGAGCGGGTCTCTACTCTGTATAATTCCTTAT 1664
 QY 629 uYrlylMetSerTrrLysAspPheProGlnLeu-----PheAsnGlyGlyGlnAl 646
 Db 1665 GTATAAATGCTACTACAGATTGAGAAATGAGCTAGTATTTGCACACCCCGG 1724
 QY 646 aThrAspArgValArgGlnMetIleThrProLeuAspValProLeuAspTrrPh 666
 Db 1725 TTTTACCCAGACACGTAATGCTGAGATCGGAATAAGGACATT---AAATTCAAGCATTT 1781
 QY 666 eAspGluValPheThrSerGluAsnTrrMetValArgIleTrrGlnLeuLysLysAspAs 686
 Db 1782 CGAGAGAGCCTTATACATGAGCACTGGCTGTAGGATATATAAGTGAAGACCCGA 1841
 QY 686 p 686
 Db 1842 C 1842

RESULT 15
 CNS06YOT 981 bp DNA linear GSS 06-JUL-2001
 LOCUS T3 end of clone AY0AA05C11 of library AY0AA from strain CBS 6340
 DEFINITION Kluyveromyces thermotolerans, genomic survey sequence.
 ACCESSION AL420435
 VERSION AL420435.1 GI:12203620
 KEYWORDS GSS.
 SOURCE Kluyveromyces thermotolerans
 ORGANISM Kluyveromyces thermotolerans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 1 (bases 1 to 981)
 Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
 Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
 de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
 Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
 Saurin, W., Tekais, P., Toffano-Nioche, C., Wesolowski-Louvel, M.,
 Wincker, P., and Weissenbach, J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)
 20594711
 11152876
 2 (bases 1 to 981)
 Malpertuy, A., Llorente, B., Blandin, G., Artiguenave, F., Wincker, P.,
 and Dujon, B.
 Genomic exploration of the hemiascomycetous yeasts: 10.
 Kluyveromyces thermotolerans
 FEBS Lett. 487 (1), 61-65 (2000)
 20594720
 11152885
 3 (bases 1 to 981)
 Genoscope.
 Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source
 1. 981
 /organism="Kluyveromyces thermotolerans"
 /mol_type="genomic DNA"
 /strain="CBS 6340"
 /db_xref="taxon:4916"
 /clone="AY0AA05C11"
 /clone_lid="AY0AA"
 /note="Tend : T3"
 <7..>939
 misc_feature
 /note="similar to Saccharomyces cerevisiae ORF YGL022w [S113 ; oligosaccharyl transferase subunit]"
 /evidence=not_experimental

ORIGIN

Alignment Scores:
 Pred. No.: 1,3e-134 Length: 981
 Score: 1303.50 Matches: 250
 Percent Similarity: 84.71% Conservative: 27
 Best Local Similarity: 76.43% Mismatches: 41
 Query Match: 34.66% Indels: 10
 DB: 9 Gaps: 2

US-10-028-384-6 (1-718) x CNS06YOT (1-981)

QY 336 AsnTrrAlaLysIleHisIleProIleAlaSerValSerGluHisGlnProValSer 355
 Db 1 AACACAGGAGATCCACATCCCATTTATCGCTCGGTATCAGACACGACCTACCGCG 60
 QY 356 TrrProAlaPhePhePheAspThrHisPheLeuIleTrrPheProAlaGlyValPhe 375
 Db 61 TGGCCAGCCTTTCTTTGTATCTACTCAATTTCTGATTTGGTTATTCACGCGGGTATTT 120
 QY 376 LeuLeuPheLeuAspLeuHisValPheValIleAlaTrrSerValLeuCys 395
 Db 121 TACTGTTCTTGGATCTCAAGATGAACACGTTTCTGATTTCTTACTCTCGTTCTCTGC 180
 QY 396 SerTrrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValIleCysVal 415
 Db 181 TCTTATTTCGGGTGTCTGTCAGACTCATGCTAATCTGACACCAATCATCTGTGTC 240
 QY 416 SerAlaAlaValAlaLeuSerLysIlePheAspIleTrrLeuAspPheLys---ThrSer 434
 Db 241 TGTGCGGAGTAGGCATTTCCAGCTATTCGAGCTCTACGCCGACTTTTCATATTTTTC 300
 QY 435 AspArgLysTrrAlaIleLysPro-----AlaAlaLeuLeu 446
 Db 301 AGTTCCTCCAGCGCAGAAAGCCAAACCTGTCCAAGAGACGCCGATTATGACTTTGGTT 360
 QY 447 AlalysLeuIleValSerGlySerPheIlePheTrrLeuTrrLeuPheValPheHisSer 466
 Db 361 ACTAAGCTAACGGTTTGTGGTCTTCTTCTGGCCCTATTGTGTTGTGCGACCACTGT 420
 QY 467 ThrTrrPValThrArgTrrAlaTrrSerSerProSerValValLeuProSerGlnTrrPro 486
 Db 421 ACCTGGGTGACCTCTAACGCTTATTCTGCGCCATTCGTGGTGTTCCTTCACGTAATCC 480
 QY 487 AspGlyLysLeuAlaLeuIleAspPheArgGluAlaTrrTrrPheLeuArgMetAsn 506
 Db 481 GACGATCTCTGCTCTGATAGACGACTTCGGTAGGCTTACTACTGGGTGGCTATGAAC 540
 QY 507 SerAspGluAspSerLysValAlaAlaTrrTrrAspTrrGlyTrrGlnIleGlyGlyMet 526
 Db 541 ACTGCAGAGATGCCAGGTTCCTCTTGTGGGACTACGGCTACCAATTTGTGGCATG 600
 QY 527 AlaAspArgThrTrrLeuValAspAsnThrTrrAsnAsnThrHisIleAlaIleVal 546
 Db 601 GCTGACGAACACCTGGTAGACAACAATCTTGGAAACCAACCCACCATCGTATCGTG 660

```
QY 547 GlyLysAlaMetAlaSerProGluGluLysSerTyrGluIleLeuLysGluHisAspVal 566
Db 661 GGTAAGCTATGTCGTCCTCCCAAGAGAGAGCGGTATGACATCTTGAAGGCTCAGCATGTG 720
QY 557 AspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGlyGlyAspAspIleAsnLys 586
Db 721 GACTACGTTCTGGTTATTTTCGGCGGCTCATAGGATTCACGGTATGATATCAACAAG 780
QY 587 PheLeuTyrMetIleArgIleSerGluGlyIleTyrProGluGluIleLysGluArgTyr 606
Db 781 TTCCTATGGATGGTAAGGATTTCTGAGGCGCATATGCCCGCATGAATCAAGAGAGAGAC 840
QY 607 PheTyrThrAlaGluGlyGlyTyrArgValAspAlaArgAlaSerGluThrMetArgAsn 626
Db 841 TACTTCACCCCGGCGGTGAGTACCGGTATGGCGCGCAACTDTCCCAACCATGAAGAC 900
QY 627 SerLeuLeuTyrLysMetSerTyrLysAspPheProGluLeuPheAsnGlyGlyGlnAla 646
Db 901 TCGCTGTGTGTACAGCTCTCTACCCGCTTCCGACA-TTGTTCATAGGSGTTGAGGCG 959
QY 647 ThrAspArgValArgGlnGln 653
Db 960 ATCGACCGTCTGCGTGTCTAG 980
```

Search completed: December 15, 2004, 11:51:40
Job time : 5663.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 02:41:23 ; Search time 16231.6 Seconds
(without alignments)
7041.770 Million cell updates/sec

Title: US-10-028-384-7
Perfect score: 2417
Sequence: 1 tctaaagcgaagaatgtgctg.....ccaaaaaataaaaaaa 2417

Scoring table: OLIGO.NUC
Gapop 50.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgs.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_scs.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2417	100.0	2417	3	AF132552 Drosophil
2	2417	100.0	2417	6	AX799088 Sequence
3	2409	99.7	2899	6	CQ589353 Sequence
C 4	1463	60.5	4922	6	CQ589352 Sequence
C 5	1463	60.5	162321	3	AC007853 Drosophil
C 6	1463	60.5	181132	3	AC008206 Drosophil
C 7	1463	60.5	227219	3	AE003750 Drosophil
C 8	1342	55.5	75650	2	AC018145 Drosophil
C 9	696	28.8	2953	6	CQ596819 Sequence
C 10	249	10.3	2785	6	CQ596795 Sequence
11	29	1.2	297	8	AF548377 Talaromyc
12	29	1.2	218635	2	AC118338 Rattus no
13	29	1.2	263377	2	AC105647 Rattus no
14	29	1.2	285045	2	AC116217 Rattus no
15	28	1.2	1968	9	AK074363 Homo sapi
16	28	1.2	110000	2	PFMAL13.26 Continuation (27 o
C 17	28	1.2	196966	10	AC087898 Mus muscu
C 18	28	1.2	268583	2	AC111344 Rattus no
C 19	27	1.1	391	6	CQ433488 Sequence

C 20	27	1.1	402	6	CQ415729 Sequence
C 21	27	1.1	471	6	CQ424280 Sequence
C 22	27	1.1	1420	3	AY513651 Ostrinia
C 23	27	1.1	1420	3	AY513652 Ostrinia
C 24	27	1.1	2158	5	BC061422 Xenopus t
C 25	27	1.1	2270	9	BC041156 Homo sapi
C 26	27	1.1	2474	9	BSM806789 Homo sapi
C 27	27	1.1	156070	2	AC110726 Mus muscu
C 28	27	1.1	173871	2	AC102469 Mus muscu
C 29	27	1.1	180998	2	AC113766 Mus muscu
C 30	27	1.1	185342	2	CR387989 Dario rer
C 31	26	1.1	786	10	BC037513 Mus muscu
C 32	26	1.1	1332	8	AF370133 Arabidops
C 33	26	1.1	1888	3	AX112753 Clona int
C 34	26	1.1	2855	6	CQ573675 Sequence
C 35	26	1.1	3248	6	AX753657 Sequence
C 36	26	1.1	3418	9	BC037539 Homo sapi
C 37	26	1.1	4822	10	BC064459 Mus muscu
C 38	26	1.1	6153	6	CQ573674 Sequence
C 39	26	1.1	9765	3	AY160106 Dictyoste
C 40	26	1.1	18405	2	AC017914 Drosophil
C 41	26	1.1	104992	2	AC005504 Plasmodiu
C 42	26	1.1	110000	8	CR382128 21 Continuation (22 o
C 43	26	1.1	114136	10	AL586447 Mouse DNA
C 44	26	1.1	161577	2	AC118924 Mus muscu
C 45	26	1.1	166908	3	AC104631 Drosophil

ALIGNMENTS

RESULT 1	AF132552	Drosophila melanogaster	2417 bp	mrna	linear	INV 12-FEB-2003
LOCUS	AF132552	Drosophila melanogaster	GM01838	full insert	cdna.	
DEFINITION	AF132552	Drosophila melanogaster	GM01838	full insert	cdna.	
ACCESSION	AF132552	Drosophila melanogaster	GM01838	full insert	cdna.	
VERSION	FLI CDNA.	Drosophila melanogaster	GM01838	full insert	cdna.	
KEYWORDS	FLI CDNA.	Drosophila melanogaster	GM01838	full insert	cdna.	
SOURCE	FLI CDNA.	Drosophila melanogaster	GM01838	full insert	cdna.	
ORGANISM	FLI CDNA.	Drosophila melanogaster	GM01838	full insert	cdna.	
REFERENCE	1	(bases 1 to 2417)				
AUTHORS	Tsang, G., Brokstein, P., Frise, E., Harvey, D., Evans-Holm, M., Lewis, S.E., Suh, C. and Rubin, G.M.					
TITLE	Direct Submission					
JOURNAL	Submitted (02-MAR-1999) Berkeley Drosophila Genome Project, University of California Berkeley, 539 Life Sciences Addition #3200, Berkeley, CA 94720, USA					
REFERENCE	2	(bases 1 to 2417)				
AUTHORS	Stapleton, M., Brokstein, P., Hong, L., Agbavani, A., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Kronmiller, B., Li, P., Liao, G., Miranda, A., Mungall, C.J., Nuncio, J., Pacleb, J., Pargass, V., Park, S., Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M. and Celniker, S.					
TITLE	Direct Submission					
JOURNAL	Submitted (12-FEB-2003) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA					
COMMENT	Sequence submitted by: Lawrence Berkeley National Laboratory					
	Berkeley Drosophila Genome Project					
	Lawrence Berkeley National Laboratory					
	Berkeley, CA 94720					
	This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal					

priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES

Source
1..2417
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
<1..2417
/gene="OstStt3"
/db_xref="FLYBASE:FBgn0011336"
<1..2395
/gene="OstStt3"
/note="Longest ORF"
/codon_start=2
/product="GM01838p"
/protein_id="AAD27851.2"
/db_xref="GI:28317200"
/db_xref="FLYBASE:FBgn0011336"
/translation="LSRECVAFQIGNPVNTGWNMDNRTPKMLNSKVAGYSSLLTF
ALLIALLAGFSLFAVIREPSILHEEPFNXYRATAMVQNGWYNPLNPFDEBAY
ELWAGSGTVPGMLTSGGIHMLLVNIPVHIDICVFLAPISGTSSTLYLTK
SAAALSYFNVMSAWGYFIINILPHVFLIMRGYSRPLLTYSYFLLGLFSM
QIPVGPQPTIRSEHMAALGVFLMLAVLRLHQLSVLRNEFRKLFYVGLLVGV
FVAVVITMLGVAPWSGRFYLWDGTAKIHPIIASVSEHQPWTWFFFDHLILV
CAFVGVYCIKQINDREVFVLYAISAVFAGVWVLMILTTPVCMAGVAFSGLL
LVLEQSSKEMGTASAEVDSAEDESLEKTLXDKAKLKHRTKHDAQDDTVGSN
LKSIVLAVLMLMFAVHCTWTSNAYSPSIVLAFHNSODGSNLLIDREBAYML
SONTADARWMSWDYQYIAGMANRTILVDNNWNNSHALVGMKAMSTSEKSEIM
ISLDVDFVLVFGVIGVSGDINKLFWIRIAGBHPKDIKESDYFDLEGEFRVDAE
GAPALLNKLKYSYFBEKLDYRPGSYDRNNAVIGNKDFDLILEEAYFTEHW
LVRIYRVKPFHEFNRPDLKERTIPANFISFKNSRRKRYIRNRPVVVRKRLTK"

ORIGIN

Query Match 100.0%; Score 2417; DB 3; Length 2417;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAAGCGAAGATGTGCTGATTCAGATCGGTTATATTTTCAGATGATCGGCTG 60
DB 1 TCTAAGCGAAGATGTGCTGATTCAGATCGGTTATATTTTCAGATGATCGGCTG 60
QY 61 GAATGGGACATGAATCGAGCGCGAAGATGCTGAACAGCAAGTGGCTACAGAG 120
DB 61 GAATGGGACATGAATCGAGCGCGAAGATGCTGAACAGCAAGTGGCTACAGAG 120
QY 121 CCTAATCATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 121 CCTAATCATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 181 CGCGCTCATCGCTTCAGTGCATATTCATGAGTTGATCGGTGCTCAATACCGGC 240
DB 181 CGCGCTCATCGCTTCAGTGCATATTCATGAGTTGATCGGTGCTCAATACCGGC 240
QY 241 CACCGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
DB 241 CACCGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 301 ATGGTATCGCTCGCGAGGATGTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCT 360
DB 301 ATGGTATCGCTCGCGAGGATGTGGCGGTGCTGCTGCTGCTGCTGCTGCT 360
QY 361 CGCGGGAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 361 CGCGGGAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 421 COTGTTCCTGGCGCGGATCTTCAGTGGCCTGACCTCCATCCACCTACCTGCTGACCAA 480

DB 421 CGTGTTCCTGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480
QY 481 GAGAGCTGTGGTTCGGCGGCGCGGCTCTTTCGCGCGCAGCTTTCATCGCCTGCTGG 540
DB 481 GAGAGCTGTGGTTCGGCGGCGCGGCTCTTTCGCGCGCAGCTTTCATCGCCTGCTGG 540
QY 541 CTACATCAGTAGTGGTGGCTGGATCGTACGATAAGAGGGCATTCGCATATTCGCGCT 600
DB 541 CTACATCAGTAGTGGTGGCTGGATCGTACGATAAGAGGGCATTCGCATATTCGCGCT 600
QY 601 GAGAGCTACCTACTTCTGCTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 601 GAGAGCTACCTACTTCTGCTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
QY 661 CGCAGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
DB 661 CGCAGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 721 CAACCTGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 721 CAACCTGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 781 GCTGACAGCTACAGCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 781 GCTGACAGCTACAGCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY 841 CGTGGAGATTCACCGATACGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 841 CGTGGAGATTCACCGATACGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
QY 901 CCTATATGGCGTGGCGACCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB 901 CCTATATGGCGTGGCGACCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 961 GAGAGCTGTTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 961 GAGAGCTGTTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
QY 1021 GCTCAGCATGCTGGCGCTTGTGGCGCGTGGAGTGGAGCTTCTACTCGCTGCTGGAGTAC 1080
DB 1021 GCTCAGCATGCTGGCGCTTGTGGCGCGTGGAGTGGAGCTTCTACTCGCTGCTGGAGTAC 1080
QY 1081 TGCTAGCGAGATTCACATTCCTCATCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 1081 TGCTAGCGAGATTCACATTCCTCATCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 1141 TTGCTTCTGCTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1141 TTGCTTCTGCTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
QY 1201 GATCTGCTACAGCAGATCAAGCAGCGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
DB 1201 GATCTGCTACAGCAGATCAAGCAGCGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
QY 1261 GGTTTACTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
DB 1261 GGTTTACTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
QY 1321 GCTGGCGGAGTGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 GCTGGCGGAGTGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
QY 1381 GCGAATGGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
DB 1381 GCGAATGGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
QY 1441 GAGAGAGCTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1500
DB 1441 GAGAGAGCTGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1500
QY 1501 GAGAGATCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
DB 1501 GAGAGATCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560

QY 1561 GTTGATGATGTTGGTGTCTTCCAGACAGTGGGTGACCAAGTCTACTCCAGTCCCTC 1620
Db 1561 GTTGATGATGTTGGTGTCTTCCAGACAGTGGGTGACCAAGTCTACTCCAGTCCCTC 1620
QY 1621 CATTTGTTGGCTTTCCCAACAGTCAAGATGGATCCCGCAACATTTTAGAGCATTTAG 1680
Db 1621 CATTTGTTGGCTTTCCCAACAGTCAAGATGGATCCCGCAACATTTTAGAGCATTTAG 1680
QY 1681 AGAGGCTTACTACTGGCTTTCCAGAACACTGCCGATGATGCTCGCGTTATGTTGGTG 1740
Db 1681 AGAGGCTTACTACTGGCTTTCCAGAACACTGCCGATGATGCTCGCGTTATGTTGGTG 1740
QY 1741 GGATTTACGGATACAGATAGCGGGAATGGCAAAACAGAACGACGCTAGTGGATAATAATAC 1800
Db 1741 GGATTTACGGATACAGATAGCGGGAATGGCAAAACAGAACGACGCTAGTGGATAATAATAC 1800
QY 1801 GTGGAACAAATAGTACATAGGCTGGTTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1860
Db 1801 GTGGAACAAATAGTACATAGGCTGGTTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1860
QY 1861 CTACGAATTTATGACATCTCTTGGACGTGGAATGCTTGGTGTATCTTTGGCGGTGTAT 1920
Db 1861 CTACGAATTTATGACATCTCTTGGACGTGGAATGCTTGGTGTATCTTTGGCGGTGTAT 1920
QY 1921 CGGCTATTCTGGCGATGATATCAACAAGTCTCTGGATGCTGCGAATTTCTGAGGAGA 1980
Db 1921 CGGCTATTCTGGCGATGATATCAACAAGTCTCTGGATGCTGCGAATTTCTGAGGAGA 1980
QY 1981 GCATCCCAAGGACATTAAGGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGA 2040
Db 1981 GCATCCCAAGGACATTAAGGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGA 2040
QY 2041 TGCCGAAGGTGCTCGGCGCTGCTCAACTGCTTATGTACAAATTAAGTACTACAGATT 2100
Db 2041 TGCCGAAGGTGCTCGGCGCTGCTCAACTGCTTATGTACAAATTAAGTACTACAGATT 2100
QY 2101 CGGGAATTTGAAGTTGGACTACAGAGTCCATCTGGATATGATCGACAGTAAACGCGGT 2160
Db 2101 CGGGAATTTGAAGTTGGACTACAGAGTCCATCTGGATATGATCGACAGTAAACGCGGT 2160
QY 2161 CATTTGGGAATTAAGACTTGGATCTGACCTACCTGAGAGGCGCTTACACCAACACACTG 2220
Db 2161 CATTTGGGAATTAAGACTTGGATCTGACCTACCTGAGAGGCGCTTACACCAACACACTG 2220
QY 2221 GCTTTGTCGATCTATAGGTTGAAGAGCGCATGAGTTCAATAGACCATCTAGAGAC 2280
Db 2221 GCTTTGTCGATCTATAGGTTGAAGAGCGCATGAGTTCAATAGACCATCTAGAGAC 2280
QY 2281 CAAGGAGAGAACGATTCCTCCAGAAACTTCAATTCGAGAAAGAACTTAAAGCGTCCGAA 2340
Db 2281 CAAGGAGAGAACGATTCCTCCAGAAACTTCAATTCGAGAAAGAACTTAAAGCGTCCGAA 2340
QY 2341 GGGCTACATACGAAACCGAGCGGTGTTGTTAAGGGAACGAACTTGAATTAACCCCA 2400
Db 2341 GGGCTACATACGAAACCGAGCGGTGTTGTTAAGGGAACGAACTTGAATTAACCCCA 2400
QY 2401 AAAAAAAAAAAAAAAAAA 2417
Db 2401 AAAAAAAAAAAAAAAAAA 2417

RESULT 2
AX799088 2417 bp DNA linear PAT 08-OCT-2003
LOCUS
DEFINITION
Sequence 7 from Patent WO03054008.
ACCESSION
AX799088
VERSION
AX799088.1 GI:37605061
KEYWORDS
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1
Perrault, C. and McBride, K.
Mammalian SIMP protein, gene sequence and uses thereof in cancer
therapy
Patent: WO 03054008-A 7 03-JUL-2003;
Comptigene Inc. (CA)

Location/Qualifiers
1. 2417

source

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/note="STT3, Acc# AF132552"

ORIGIN

Query Match 100.0%; Score 2417; DB 6; Length 2417;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAAGCGAAGATGTGCTGTTCGATTTCAGATTCGGTTATATTTTCGAGTTACTGGCTG 60

Db 1 TCTAAGCGAAGATGTGCTGTTCGATTTCAGATTCGGTTATATTTTCGAGTTACTGGCTG 60

QY 61 GAATTTGGGACATGAATCGGACCGGAGATGCTGAACAGCAAGGTGGTGGCTACAGCAG 120

Db 61 GAATTTGGGACATGAATCGGACCGGAGATGCTGAACAGCAAGGTGGTGGCTACAGCAG 120

QY 121 CTTAATCACCTTCGCGCATCTCTAATCGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 180

Db 121 CTTAATCACCTTCGCGCATCTCTAATCGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 180

QY 181 CGCGGTATTCGCTTCGAGTTCGATTCATCATGAGTTTGCCTGCTGCTGCTGCTGCTGCTGCT 240

Db 181 CGCGGTATTCGCTTCGAGTTCGATTCATCATGAGTTTGCCTGCTGCTGCTGCTGCTGCTGCT 240

QY 241 CACCGCTACATGCTGCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300

Db 241 CACCGCTACATGCTGCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300

QY 301 ATGGTATTCGCTTCGCGAGATTTGGGGGGTACCGCTCTATCCCGGCTGATGATTACGCTC 360

Db 301 ATGGTATTCGCTTCGCGAGATTTGGGGGGTACCGCTCTATCCCGGCTGATGATTACGCTC 360

QY 361 CGCGCGAATCCATTTGGCTGCTGCGAGTACTCAATACCGGTTCATATTCGTCGATCATCTG 420

Db 361 CGCGCGAATCCATTTGGCTGCTGCGAGTACTCAATACCGGTTCATATTCGTCGATCATCTG 420

QY 421 CGTGTTCCTGCGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480

Db 421 CGTGTTCCTGCGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480

QY 481 GGAGTGTGCTTCGCGCGCGCGCGCTCTTCGCGCGAGTTCATTCGCGCTGCTGCTGCTGCTG 540

Db 481 GGAGTGTGCTTCGCGCGCGCGCGCTCTTCGCGCGAGTTCATTCGCGCTGCTGCTGCTGCTG 540

QY 541 CTACATCAGTAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600

Db 541 CTACATCAGTAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600

QY 601 GCAGTTACCTTCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660

Db 601 GCAGTTACCTTCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660

QY 661 CGCAGCGCTTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 720

Db 661 CGCAGCGCTTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 720

QY 721 CAACCTGATACCTTCGACGCTCTTCGTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780

Db 721 CAACCTGATACCTTCGACGCTCTTCGTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780

QY 781 GCTGACGAGCTTACAGACCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840

Db 781 GCTGACGAGCTTACAGACCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840

Db 781 GCTGACCACTACAGACCTTCTACATCCTGGAGTGTCTTCTCCATGCGAGATCCCGTT 840
 Qy 841 CGTGGGATTCACAGATACGACAGTGAACACATGCTGCTGGAGTGTGTGCT 900
 Db 841 CGTGGGATTCACAGATACGACAGTGAACACATGCTGCTGGAGTGTGTGCT 900
 Qy 901 CTTATGCGGTGGCCACTTTCGGCAATTTGAGTCCGTCGTGTCGCGCAACAGATTCG 960
 Db 901 CTTATGCGGTGGCCACTTTCGGCAATTTGAGTCCGTCGTGTCGCGCAACAGATTCG 960
 Qy 961 GAAGCTGTTATCCTGCGGGAATGCTGTCGCGGATTTGCGTCTTGTGGCGCTGCTG 1020
 Db 961 GAAGCTGTTATCCTGCGGGAATGCTGTCGCGGATTTGCGTCTTGTGGCGCTGCTG 1020
 Qy 1021 GCTCACCATGCTGGCGTGTGGCCCGCTGGAGTGGAGCTTCTACTCGCTGTGGATAC 1080
 Db 1021 GCTCACCATGCTGGCGTGTGGCCCGCTGGAGTGGAGCTTCTACTCGCTGTGGATAC 1080
 Qy 1081 TGCTAGCGCAAGATCCATTCCTCCATCATTCGATCCGTCGAGCATCAGCCACCAC 1140
 Db 1081 TGCTAGCGCAAGATCCATTCCTCCATCATTCGATCCGTCGAGCATCAGCCACCAC 1140
 Qy 1141 TTGTTCTGCTTCTTCTTGTGATCTGCACATCTGCTGTCGCGCTTCCAGTGGAGTGTG 1200
 Db 1141 TTGTTCTGCTTCTTCTTGTGATCTGCACATCTGCTGTCGCGCTTCCAGTGGAGTGTG 1200
 Qy 1201 GTACTGATCAAGCAGATCAACAGCAGCGGTTTCGTGCTGCTGAGCCATCAGTGC 1260
 Db 1201 GTACTGATCAAGCAGATCAACAGCAGCGGTTTCGTGCTGCTGAGCCATCAGTGC 1260
 Qy 1261 GGTATTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 Db 1261 GGTATTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 Qy 1321 GCTGGCGGAGTGGCTTTTCGGGACTGTTGATGTTGATGTTGATGTTGATGTTGATGTT 1380
 Db 1321 GCTGGCGGAGTGGCTTTTCGGGACTGTTGATGTTGATGTTGATGTTGATGTTGATGTT 1380
 Qy 1381 GCGAATGGGACAGCAGATAGCCAGCCAGCCAGGAGTGAAGCTGAGATTCATTCGA 1440
 Db 1381 GCGAATGGGACAGCAGATAGCCAGCCAGGAGTGAAGCTGAGATTCATTCGA 1440
 Qy 1441 GAAGAAGACGCTGTACGACAGGCTGCGAAGCTGGAAGCATGCTACTAAGCATGATCCCA 1500
 Db 1441 GAAGAAGACGCTGTACGACAGGCTGCGAAGCTGGAAGCATGCTACTAAGCATGATCCCA 1500
 Qy 1501 GCAGGATCTGGGTCAGCTCCAACTGAGAGTATGTTATTTGGCGCTTCTAATGCT 1560
 Db 1501 GCAGGATCTGGGTCAGCTCCAACTGAGAGTATGTTATTTGGCGCTTCTAATGCT 1560
 Qy 1561 GTTCATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Db 1561 GTTCATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Qy 1621 CATTTGCTGCTTTCACACAGTCAAGTGAAGTCCCGCAACATTTAGACGATTCAG 1680
 Db 1621 CATTTGCTGCTTTCACACAGTCAAGTGAAGTCCCGCAACATTTAGACGATTCAG 1680
 Qy 1681 AGAGGCTTACTGCTGCTTTCGAGAACACTGCGGATGATGCTCGGCTTATGCTTGGTG 1740
 Db 1681 AGAGGCTTACTGCTGCTTTCGAGAACACTGCGGATGATGCTCGGCTTATGCTTGGTG 1740
 Qy 1741 GGAATACGATACAGATAGCGGAATGGGAAACAGAACAGCAGTGTGATTAATATAC 1800
 Db 1741 GGAATACGATACAGATAGCGGAATGGGAAACAGAACAGCAGTGTGATTAATATAC 1800
 Qy 1801 GTGGACATAGTACATAGCGCTGTTGGCAGGCAATGCTTCAACCGAGAGAGTC 1860
 Db 1801 GTGGACATAGTACATAGCGCTGTTGGCAGGCAATGCTTCAACCGAGAGAGTC 1860
 Qy 1861 CTACGAATATGACATCTTTGACGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 Db 1861 CTACGAATATGACATCTTTGACGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920

Qy 1921 CGGCTATTTCTGGGATGATATCAACAGTTCCTGCTGGATGGTTCGAAATGCTGAGGGAGA 1980
 Db 1921 CGGCTATTTCTGGGATGATATCAACAGTTCCTGCTGGATGGTTCGAAATGCTGAGGGAGA 1980
 Qy 1981 GCATCCCAAGGACATTAAGGAAAGGATTAATTTTACCAGCCGCGTGAATTCAGGGTAGA 2040
 Db 1981 GCATCCCAAGGACATTAAGGAAAGGATTAATTTTACCAGCCGCGTGAATTCAGGGTAGA 2040
 Qy 2041 TGCCGAAGTGTCTCGGCGCTGCTCACTGCTTATGTACAAATTAAGCTACTACAGATT 2100
 Db 2041 TGCCGAAGTGTCTCGGCGCTGCTCACTGCTTATGTACAAATTAAGCTACTACAGATT 2100
 Qy 2101 CGGGGAATTAAGTTCGACTACAGAGTCCATCTCGATATGATCGCACACGCTAACCGCT 2160
 Db 2101 CGGGGAATTAAGTTCGACTACAGAGTCCATCTCGATATGATCGCACACGCTAACCGCT 2160
 Qy 2161 CATTTGGAATAAGGACTTCGATCTGACCTACCTGGAGGAGGCTTACACACAGACACTG 2220
 Db 2161 CATTTGGAATAAGGACTTCGATCTGACCTACCTGGAGGAGGCTTACACACAGACACTG 2220
 Qy 2221 GCTTGTTCGATCATATAGGAGTGAAGCGCGATGATTTCAATAGACCATCTACTGAGAC 2280
 Db 2221 GCTTGTTCGATCATATAGGAGTGAAGCGCGATGATTTCAATAGACCATCTACTGAGAC 2280
 Qy 2281 CAAGGAGAGAACGATTCCTCCAGCAAACTTCAATTCGAGAAAGAACTTAAGGCTGCA 2340
 Db 2281 CAAGGAGAGAACGATTCCTCCAGCAAACTTCAATTCGAGAAAGAACTTAAGGCTGCA 2340
 Qy 2341 GGGCTACATACGAAACCGACCGCTTGTGTTAAGGAAACGAACTTGAATTAACCCA 2400
 Db 2341 GGGCTACATACGAAACCGACCGCTTGTGTTAAGGAAACGAACTTGAATTAACCCA 2400
 Qy 2401 AAAAAAAAAAAAAA 2417
 Db 2401 AAAAAAAAAAAAAA 2417

RESULT 3
 CO589353 2699 bp DNA linear PAT 02-FEB-2004
 LOCUS Sequence 17111 from Patent WO0171042.
 DEFINITION CO589353
 ACCESSION CO589353
 VERSION CO589353.1 GI:41648215
 KEYWORDS
 SOURCE Drosophila sp.
 ORGANISM Drosophila sp.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1
 Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
 Detection kits, such as nucleic acid arrays, for detecting the
 expression of 10,000 or more Drosophila genes and uses thereof
 Patent: WO 0171042-A 17111 27-SEP-2001;
 PE Corporation (NY) (US)
 JOURNAL Location/Qualifiers
 FEATURES
 source
 1..2699
 /organism="Drosophila sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:7242"

ORIGIN
 Query Match 99.7%; Score 2409; DB 6; Length 2699;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TCTAGCGAAGATGTGTGCTTTCGATTTTCAGATCGTGTATTAATTTTCAGTTACTGCTG 60
 Db 10 TCTAGCGAAGATGTGTGCTTTCGATTTTCAGATCGTGTATTAATTTTCAGTTACTGCTG 69
 Qy 61 GAATTTGGACATGATCGGAGCGGAGATGCTGAACAGCAAGTGGTGGCTGCTACAGCAG 120

Db 70 GAATTGGACATGAATCGGACGCGGAGATGCTGAACAGCAAGGTGGCTGCTACAGCAG 129
QY 121 CTTAATCACTTCGCGCATCTGCTAATCGCTGCTGCGCGGATTTTCCTCTCGCCCTTT 180
Db 130 CTTAATCACTTCGCGCATCTGCTAATCGCTGCTGCGCGGATTTTCCTCTCGCCCTTT 189
QY 181 CGCGCTCATCGTTTCGAGTGAATATCATGAGTTTGAATCCGCTGCTCAACTACCGGC 240
Db 190 CGCGCTCATCGTTTCGAGTGAATATCATGAGTTTGAATCCGCTGCTCAACTACCGGC 249
QY 241 CACCGCTACATGCTGCGAGATGTTTGTGTACAACTTCTCAACTGCTTCCAGCAGCGGC 300
Db 250 CACCGCTACATGCTGCGAGATGTTTGTGTACAACTTCTCAACTGCTTCCAGCAGCGGC 309
QY 301 ATGGTATCGCTCGCGCAGATGCTGCGGCTGACGCTATATCCCGGCTGATGATACGTC 360
Db 310 ATGGTATCGCTCGCGCAGATGCTGCGGCTGACGCTATATCCCGGCTGATGATACGTC 369
QY 361 CGCGGGAATCCATTGCTGCTGCGAGTACTCAACTACCGGCTCATATTCGTCACATCTG 420
Db 370 CGCGGGAATCCATTGCTGCTGCGAGTACTCAACTACCGGCTCATATTCGTCACATCTG 429
QY 421 CGTGTCTCGCGCGCGATCTTCAGTGGCTGACCTCATCTCCACCTACCTGCTGACCAA 480
Db 430 CGTGTCTCGCGCGCGATCTTCAGTGGCTGACCTCATCTCCACCTACCTGCTGACCAA 489
QY 481 GGAGTGTGGTCCGCGGCGCGGCTCTTCGCGCGCGCTTCTCCGCGCGCTTCTGCTGCG 540
Db 490 GGAGTGTGGTCCGCGGCGCGGCTCTTCGCGCGCGCTTCTCCGCGCGCTTCTGCTGCG 549
QY 541 CTACATCAGTAGTGGTGGCTGATCGTACGNTAAAGAGGATTCGCTATTCGCTTCCGCT 600
Db 550 CTACATCAGTAGTGGTGGCTGATCGTACGNTAAAGAGGATTCGCTATTCGCTTCCGCT 609
QY 601 GCAGTTCACCTACTTCTGCTGGTGGCTCAGTGAAGACTGGATCCGCTTCTGCTGCGC 660
Db 610 GCAGTTCACCTACTTCTGCTGGTGGCTCAGTGAAGACTGGATCCGCTTCTGCTGCGC 669
QY 661 CGCAGCGCTTGTCTTACTTCTACATGGTGTGCGCTGGGCTGAGTGTTCATCAT 720
Db 670 CGCAGCGCTTGTCTTACTTCTACATGGTGTGCGCTGGGCTGAGTGTTCATCAT 729
QY 721 CAACCTGATACCCCTGACGCTTCTGCTACTGCTCATATGCGCAGGTAATCGCGGCTCT 780
Db 730 CAACCTGATACCCCTGACGCTTCTGCTACTGCTCATATGCGCAGGTAATCGCGGCTCT 789
QY 781 GCTGACAGCTACAGCACTTCTACATCTGGGCTGCTGCTTCTCATGAGATCCGCTT 840
Db 790 GCTGACAGCTACAGCACTTCTACATCTGGGCTGCTGCTTCTCATGAGATCCGCTT 849
QY 841 CGTGGGATTCACCGGATACGACCAAGTGAACACATGGCTGCGCTGGGAGTGTGCT 900
Db 850 CGTGGGATTCACCGGATACGACCAAGTGAACACATGGCTGCGCTGGGAGTGTGCT 909
QY 901 CTTATGCGCTGGGCGCTTGGCGCATTTGAGTCCGCTGCTGCTGCGGCAACGAGTTCG 960
Db 910 CTTATGCGCTGGGCGCTTGGCGCATTTGAGTCCGCTGCTGCTGCGGCAACGAGTTCG 969
QY 961 GAAGCTTTCATGCTGCGGAGTTCGCTGCTGGGCTTGGCTTCTGCTGCGGCTGGT 1020
Db 970 GAAGCTTTCATGCTGCGGAGTTCGCTGCTGGGCTTGGCTTCTGCTGCGGCTGGT 1029
QY 1021 GCTACAGCTGCGGCTTGGGCGCTTGGGCTGAGTGGAGTCTGCTGCTGCGGATC 1080
Db 1030 GCTACAGCTGCGGCTTGGGCGCTTGGGCTGAGTGGAGTCTGCTGCTGCGGATC 1089
QY 1081 TGGCTACGCCAAGATCCACATTCCTCATCTGCTGCTGCGGATCAGGCCACAC 1140
Db 1090 TGGCTACGCCAAGATCCACATTCCTCATCTGCTGCTGCGGATCAGGCCACAC 1149
QY 1141 TTGGTTCGCTTCTTCTTGTATCTGCAATCCTGGTGGCGCTTCCAGTGGAGTGTG 1200
Db 1150 TTGGTTCGCTTCTTCTTGTATCTGCAATCCTGGTGGCGCTTCCAGTGGAGTGTG 1209

QY 1201 GTACTGCATCAAGCAGATCAACGACGACGCGTTTTCGTGGTCTGTGTACGCCATCAGTGC 1260
Db 1210 GTACTGCATCAAGCAGATCAACGACGACGCGTTTTCGTGGTCTGTGTACGCCATCAGTGC 1269
QY 1261 GGTTTACTTCGCTCGTGTGATGCGTTTGTGATGTTGACCTCAGCCGCTGCTGTGCTAT 1320
Db 1270 GGTTTACTTCGCTCGTGTGATGCGTTTGTGATGTTGACCTCAGCCGCTGCTGTGCTAT 1329
QY 1321 GCTGCCGAGTGGCTTTTTCGGACTGTTGGATGTTTCTCTCAAGAGGATTCGTCTAA 1380
Db 1330 GCTGCCGAGTGGCTTTTTCGGACTGTTGGATGTTTCTCTCAAGAGGATTCGTCTAA 1389
QY 1381 GCGAATGGGACACGCGATNAGCCGACGACGAGTGGATGAGCTGAGGATTCGATTA 1440
Db 1390 GCGAATGGGACACGCGATNAGCCGACGACGAGTGGATGAGCTGAGGATTCGATTA 1449
QY 1441 GAAGAAGACGCTGTACGACAAAGCTGGCAAGCTGAAGCATCTGTAAGCATGATGCCCA 1500
Db 1450 GAAGAAGACGCTGTACGACAAAGCTGGCAAGCTGAAGCATCTGTAAGCATGATGCCCA 1509
QY 1501 GCAGTACTGGGCTGAGCTCCACCTGAAGATGATGTTTATTTGGCGCTTCTATGCT 1560
Db 1510 GCAGTACTGGGCTGAGCTCCACCTGAAGATGATGTTTATTTGGCGCTTCTATGCT 1569
QY 1561 GTTGATGATGTTGCTGCTCCACTGTCAGCTGGGTGACAGCAATGCTTCTCAGTCCCTC 1620
Db 1570 GTTGATGATGTTGCTGCTCCACTGTCAGCTGGGTGACAGCAATGCTTCTCAGTCCCTC 1629
QY 1621 CATGCTCTGCTTTCACACAGTGAAGTGGATCCGCAACATTTAGACGATTCAG 1680
Db 1630 CATGCTCTGCTTTCACACAGTGAAGTGGATCCGCAACATTTAGACGATTCAG 1689
QY 1681 AGAGCTTACTGCTGCTTTCGCAACACTGCGGATGATGCTCGGCTTATGCTTGGTG 1740
Db 1690 AGAGCTTACTGCTTTCGCAACACTGCGGATGATGCTCGGCTTATGCTTGGTG 1749
QY 1741 GGTATCGGATACAGATAGCGGAAATGGCAACAGCAACGCTAGTGAATTAATAC 1800
Db 1750 GGTATCGGATACAGATAGCGGAAATGGCAACAGCAACGCTAGTGAATTAATAC 1809
QY 1801 GTGAAACAATAGTCACATAGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1860
Db 1810 GTGAAACAATAGTCACATAGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1869
QY 1861 CTACGAATTAATGATCATCTTTCAGCTGGATAGTTCGCTATCTTTGGCGCTGAT 1920
Db 1870 CTACGAATTAATGATCATCTTTCAGCTGGATAGTTCGCTATCTTTGGCGCTGAT 1929
QY 1921 CGGCTATTTCTGGCGATGATATCAACAGTTCTCTGTGATGGTCCGAATTCGCTGAGGAGA 1980
Db 1930 CGGCTATTTCTGGCGATGATATCAACAGTTCTCTGTGATGGTCCGAATTCGCTGAGGAGA 1989
QY 1981 GCATCCGAGACATTAAGGAAGCGATTTACCGACGCGGTGAATTCAGGCTAGA 2040
Db 1990 GCATCCGAGACATTAAGGAAGCGATTTACCGACGCGGTGAATTCAGGCTAGA 2049
QY 2041 TGCCGAAGTCTCGGCGCTTCTCAACTGCTTATGTAACAAATTAAGCTACTACAGATT 2100
Db 2050 TGCCGAAGTCTCGGCGCTTCTCAACTGCTTATGTAACAAATTAAGCTACTACAGATT 2109
QY 2101 CGGGAAATTGAATTTGACTACAGAGTCCATCTGATATGATCGCACACGCTAACCGCT 2160
Db 2110 CGGGAAATTGAATTTGACTACAGAGTCCATCTGATATGATCGCACACGCTAACCGCT 2169
QY 2161 CATTTGGGAATAAGGACTTCGATCTGACCTACCTGGAGAGGCTTACACCAAGAACCTG 2220
Db 2170 CATTTGGGAATAAGGACTTCGATCTGACCTACCTGGAGAGGCTTACACCAAGAACCTG 2229
QY 2221 GCTTGTTCGCTATCTAGGTTGAAGAGCGGATGAGTTCAATAGACCATCTCTGAAGAC 2280
Db 2230 GCTTGTTCGCTATCTAGGTTGAAGAGCGGATGAGTTCAATAGACCATCTCTGAAGAC 2289

QY 2281 CAAGGAGAGACGATTCCTCCAGCAAACTTCAATTCGAGAAAGAACTCTAAGCGTCGCAA 2340
 Db 2290 CAAGGAGAGACGATTCCTCCAGCAAACTTCAATTCGAGAAAGAACTCTAAGCGTCGCAA 2349
 QY 2341 GGGCTACATACGAAACCGAGCGGTGTGTAGGGAAGAAACGACCTTGAATAAACCA 2400
 Db 2350 GGGCTACATACGAAACCGAGCGGTGTGTAGGGAAGAAACGACCTTGAATAAACCA 2409
 QY 2401 AAAAAAAA 2409
 Db 2410 AAAAAAAA 2418

RESULT 4
 CQ589352/c
 LOCUS
 DEFINITION Sequence 17110 from Patent WO0171042.
 ACCESSION CQ589352
 VERSION CQ589352.1 GI:41648214
 KEYWORDS
 SOURCE Drosophila sp.
 ORGANISM Drosophila sp.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1
 AUTHORS Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
 TITLE Detection kits, such as nucleic acid arrays, for detecting the
 expression of 10,000 or more Drosophila genes and uses thereof
 JOURNAL Patent: WO 0171042-A 17110 27-SEP-2001;
 PE Corporation (NY) (US)
 FEATURES
 source Location/Qualifiers
 1..4922
 /organism="Drosophila sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:7242"

ORIGIN
 Query Match 60.5%; Score 1463; DB 6; Length 4922;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TCTAAGCGAAGATGCTGCTTGCATTCAGATCGGTATATATTTTCGAGTTACTGGCTG 60
 Db 3913 TCTAAGCGAAGATGCTGCTTGCATTCAGATCGGTATATATTTTCGAGTTACTGGCTG 3854
 QY 61 GAATTTGGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGCGTGGCTACAGCAG 120
 Db 3853 GAATTTGGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGCGTGGCTACAGCAG 3794
 QY 121 CCTAATCACCTTCGCCATCTGCTTAATCGCTGGCTGGCGGATTTTCTCTCGCTCTT 180
 Db 3793 CCTAATCACCTTCGCCATCTGCTTAATCGCTGGCTGGCGGATTTTCTCTCGCTCTT 3734
 QY 181 CGCGCTCATCGCTTTCGAGTCGATATTCGATAGTTGATTCGCTGGCTCAATACCGGCG 240
 Db 3733 CGCGCTCATCGCTTTCGAGTCGATATTCGATAGTTGATTCGCTGGCTCAATACCGGCG 3674
 QY 241 CACCGCTTACATGCTGAGATGCTTGGTACAACTTCTCAATGCTTTCGAGCGCGCG 300
 Db 3673 CACCGCTTACATGCTGAGATGCTTGGTACAACTTCTCAATGCTTTCGAGCGCGCG 3614
 QY 301 ATGTATTCGCTCGGCGAGATTTGGCGGCTACCGTCTATTCGCGGCTGATATACGTC 360
 Db 3613 ATGTATTCGCTCGGCGAGATTTGGCGGCTACCGTCTATTCGCGGCTGATATACGTC 3554
 QY 361 CGCGGGAATCCATTTGGCTGCTGACGATCTCAACATACCGGTCCATTTCTGATGATCTG 420
 Db 3553 CGCGGGAATCCATTTGGCTGCTGACGATCTCAACATACCGGTCCATTTCTGATGATCTG 3494
 QY 421 CGTGTTCGCGCGGATCTTTCAGTGGCTGACCTCCATCTCCAGCTCTGCTGACCAA 480
 Db 3493 CGTGTTCGCGCGGATCTTTCAGTGGCTGACCTCCATCTCCAGCTCTGCTGACCAA 3434

QY 481 GGAGCTGTGGTCCGGGGCGCGGCTCTTTCGGCGCCAGCTTTCATGCCCATCTGTCCTCG 540
 Db 3433 GGAGCTGTGGTCCGGGGCGCGGCTCTTTCGGCGCCAGCTTTCATGCCCATCTGTCCTCG 3374
 QY 541 CTACATCAGTAGGTGGTGGCTGGATCGTACGATACGAGGGCATTCGCATATTCGCCCT 600
 Db 3373 CTACATCAGTAGGTGGTGGCTGGATCGTACGATACGAGGGCATTCGCATATTCGCCCT 3314
 QY 601 GCAGTTCACTACTTCTGTGGGTGGCTCAGTGAAGAGCTGGATCCGCTGTCTGTCGGC 660
 Db 3313 GCAGTTCACTACTTCTGTGGGTGGCTCAGTGAAGAGCTGGATCCGCTGTCTGTCGGC 3254
 QY 661 CGCAGCGCTTGTGCTACTTCTACATGGTTCGCGCTGGGTGGCTACGTTTCATCAT 720
 Db 3253 CGCAGCGCTTGTGCTACTTCTACATGGTTCGCGCTGGGTGGCTACGTTTCATCAT 3194
 QY 721 CAACCTGATACCCCTGCAGCTCTTCCTACTGCTCATTATGGCAGGTACTTCGCGCGCT 780
 Db 3193 CAACCTGATACCCCTGCAGCTCTTCCTACTGCTCATTATGGCAGGTACTTCGCGCGCT 3134
 QY 781 GCTGACGCTACAGCACTTCTACATCCTGGGACTGCTGTTCTCCATGAGATCCCTT 840
 Db 3133 GCTGACGCTACAGCACTTCTACATCCTGGGACTGCTGTTCTCCATGAGATCCCTT 3074
 QY 841 CGTGGGATTCGAACCGATACGCAACCAAGTGAACACATGCTGCTGGGAGTGTGTGCT 900
 Db 3073 CGTGGGATTCGAACCGATACGCAACCAAGTGAACACATGCTGCTGGGAGTGTGTGCT 3014
 QY 901 CTTATGGCGGTGGCACTTTCGCGCAATTCGAGTCCGCTGCTGCGGCAACGAGTTCG 960
 Db 3013 CTTATGGCGGTGGCACTTTCGCGCAATTCGAGTCCGCTGCTGCGGCAACGAGTTCG 2954
 QY 961 GAAGCTGTTTCATCGTCGCGGATTCGCTGGTGGGCTTGGCTCTTTCGCGCTGCTGGT 1020
 Db 2953 GAAGCTGTTTCATCGTCGCGGATTCGCTGGTGGGCTTGGCTCTTTCGCGCTGCTGGT 2894
 QY 1021 GCTCACCATGCTGGGGTGTGGCCCCGTGGAGTGGAGCTTCTACTGCTGTGGGATC 1080
 Db 2893 GCTCACCATGCTGGGGTGTGGCCCCGTGGAGTGGAGCTTCTACTGCTGTGGGATC 2834
 QY 1081 TGCTACGCAAGATCCCAATTCCTCATTCGCTGCGGAGCATCAGCCCAAC 1140
 Db 2833 TGCTACGCAAGATCCCAATTCCTCATTCGCTGCGGAGCATCAGCCCAAC 2774
 QY 1141 TTGGTTCCTGTTCTTTGATCTGCACATCTGCTGGTGGCCCTTCCAGTGGGAGTGTG 1200
 Db 2773 TTGGTTCCTGTTCTTTGATCTGCACATCTGCTGGTGGCCCTTCCAGTGGGAGTGTG 2714
 QY 1201 GTACTGCATCAAGAGATCAACGACGCGGCTTTTCGTTGCTGTACGCGCATCAGTGC 1260
 Db 2713 GTACTGCATCAAGAGATCAACGACGCGGCTTTTCGTTGCTGTACGCGCATCAGTGC 2654
 QY 1261 GGTTCATCTGCTGTGTGATGGTGGTTCGATGTTGACCTCAGCGCGGTGGTGGAT 1320
 Db 2653 GGTTCATCTGCTGTGTGATGGTGGTTCGATGTTGACCTCAGCGCGGTGGTGGAT 2594
 QY 1321 GCTGGCGGAGTGGGCTTTTCGAGTGTGTTCTGCAAGAGGATTCGTTCTAA 1380
 Db 2593 GCTGGCGGAGTGGGCTTTTCGAGTGTGTTCTGCAAGAGGATTCGTTCTAA 2534
 QY 1381 GCGAATGGGCAAGCGCATAGCGGAGCCACCGAAGTGAAGTGAAGTGAAGTGAAGTGA 1440
 Db 2533 GCGAATGGGCAAGCGCATAGCGGAGCCACCGAAGTGAAGTGAAGTGAAGTGAAGTGA 2474
 QY 1441 GAAGAGAGCGCTGTACGCAAGG 1463
 Db 2473 GAAGAGAGCGCTGTACGCAAGG 2451

RESULT 5
 AC007853/c
 LOCUS
 162921 bp DNA linear INV 27-FEB-2001

DEFINITION Drosophila melanogaster, chromosome 3R, region 96B-96C, BAC clone
 BAC0302, complete sequence.
 ACCESSION AC007853
 VERSION
 KEYWORDS HTG.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 162921)
 Celnikier, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
 Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
 Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,
 Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,
 Dodson, K., Dorset, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D.,
 Ferrier, S., Friese, E., Galle, R.F., Garg, N.S., George, R.A.,
 Gonzalez, M., Hock, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
 Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
 McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
 Pacle, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
 Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
 Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
 Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
 Sequencing of Drosophila chromosome 3R, region 96B-96C
 Unpublished
 2 (bases 1 to 162921)
 Celnikier, S.E., Agbavani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
 Butenhoff, C., Champs, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Svirskas, R., Wan, K.H., Weinburg, T., Zhang, R., Zietan, L.L. and
 Rubin, G.M.
 Direct Submission
 Submitted (17-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Feb 27, 2001 this sequence version replaced gi:5670481.
 Sequence submitted by:
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdgp@fruitfly.berkeley.edu.
 Location/Qualifiers
 1. 162921
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /map="96B-96C"
 /clone="BACR03L02 (D766)"
 /clone_lib="RPC1-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial EcoRI in
 pBACs.6)"
 Query Match 60.5%; Score 1463; DB 3; Length 162921;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TCTACGAGAGATGTCGTTGCAATTCAGATCGGTTATATATTCGAGTACTGCGTG 60
 Db 108551 TCTAACGAGAGATGTCGTTGCAATTCAGATCGGTTATATATTCGAGTACTGCGTG 108492
 QY 61 GAATTGGGACATGAATCGGACCGCGAGATGCTGAACAGAGATGCGTGGCTGCTACAGCAG 120

Db 108491 GAATTGGGACATGAATCGGACCGCGAGATGCTGAACAGAGATGCGTGGCTGCTACAGCAG 108432
 QY 121 CTAATACCTTCGCGCATCTCTGTAATCGCTGGCTGGCGGATTTCTCTCTCGCTCTT 180
 Db 108431 CTAATACCTTCGCGCATCTCTGTAATCGCTGGCTGGCGGATTTCTCTCTCGCTCTT 108372
 QY 181 CGCGCTCATCGCTTCGAGTGCATATTCATGAGTTGATCCGTTGGTTCAATACCGGC 240
 Db 108371 CGCGCTCATCGCTTCGAGTGCATATTCATGAGTTGATCCGTTGGTTCAATACCGGC 108312
 QY 241 CACCGCTACATGCTGGCAGATGGTTGTGTAACAATCTCTCAATGGTTCCAGCAGCGCG 300
 Db 108311 CACCGCTACATGCTGGCAGATGGTTGTGTAACAATCTCTCAATGGTTCCAGCAGCGCG 108252
 QY 301 ATGGTATCGCTTCGCGCAGATGTTGGCGGTACCGTCTATCCCGGCTGATGATAGTC 360
 Db 108251 ATGGTATCGCTTCGCGCAGATGTTGGCGGTACCGTCTATCCCGGCTGATGATAGTC 108192
 QY 361 CGCGCGAATCCATTCGGCTGCTGCAGTACTCAACATACCCTCCATATTCGTGACATCTG 420
 Db 108191 CGCGCGAATCCATTCGGCTGCTGCAGTACTCAACATACCCTCCATATTCGTGACATCTG 108132
 QY 421 CCGTGTCTCGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACAA 480
 Db 108131 CCGTGTCTCGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACAA 108072
 QY 481 GGAGCTGTGCTCGCGCGCGCGCTCTTCGCGCGCAGCTTCATCGCATCGTGGCTGG 540
 Db 108071 GGAGCTGTGCTCGCGCGCGCGCTCTTCGCGCGCAGCTTCATCGCATCGTGGCTGG 108012
 QY 541 CTACATCAGTAGTTCGCTGGTGGATCGTACGATAACAGAGGCAATTCGCCATATTCGCGCT 600
 Db 108011 CTACATCAGTAGTTCGCTGGTGGATCGTACGATAACAGAGGCAATTCGCCATATTCGCGCT 107952
 QY 601 CGAGTTCACCTACTTCCTGCTGGTGGCTCAGTCAAGACTGGATCCGCTGTTCTGCTCGGC 660
 Db 107951 CGAGTTCACCTACTTCCTGCTGGTGGCTCAGTCAAGACTGGATCCGCTGTTCTGCTCGGC 107892
 QY 661 CGCAGCGCTTCCTACTTCTACATAGTTCGCGCTCGCGCTGGGTGGCTGCTGTTTCATCAT 720
 Db 107891 CGCAGCGCTTCCTACTTCTACATAGTTCGCGCTCGCGCTGGGTGGCTGCTGTTTCATCAT 107832
 QY 721 CAACCTGATACCCCTGCGCAGCTTCCTGCTACTCTCATATTCGCGAGGTACTCGCGCTCT 780
 Db 107831 CAACCTGATACCCCTGCGCAGCTTCCTGCTACTCTCATATTCGCGAGGTACTCGCGCTCT 107772
 QY 781 GCTGACAGCTACAGCAGCTTCCTACATCTCGGACTGCTGTTCTCCATGAGATTCGCTT 840
 Db 107771 GCTGACAGCTACAGCAGCTTCCTACATCTCGGACTGCTGTTCTCCATGAGATTCGCTT 107712
 QY 841 CGTGGGATTCGACCGATACCGCAGTGAACACATCGCTGCGCTGGGAGTGTGTTGCT 900
 Db 107711 CGTGGGATTCGACCGATACCGCAGTGAACACATCGCTGCGCTGGGAGTGTGTTGCT 107652
 QY 901 CTTATGCGCGTGGCGCAGCTTCGCGCAGTTTCAGTCCGCTGCTGTCGCGCAACAGTTCG 960
 Db 107651 CTTATGCGCGTGGCGCAGCTTCGCGCAGTTTCAGTCCGCTGCTGTCGCGCAACAGTTCG 107592
 QY 961 GAAGCTGTTTCATCGTCGCGGATTCGCTGGGCGTGGGCTCTTCTGCGCGCTGCTGCT 1020
 Db 107591 GAAGCTGTTTCATCGTCGCGGATTCGCTGGGCGTGGGCTCTTCTGCGCGCTGCTGCT 107532
 QY 1021 GCTCACCATGCTGGCGGTTGTGGCGCGGTGGAGTGGACGCTTCTACTCGCTGTGGGATAC 1080
 Db 107531 GCTCACCATGCTGGCGGTTGTGGCGCGGTGGAGTGGACGCTTCTACTCGCTGTGGGATAC 107472
 QY 1081 TGGCTACCGCAGATCCCATTCCTCCATTCGATCGTTCGCGAGCATCAGCCAC 1140
 Db 107471 TGGCTACCGCAGATCCCATTCCTCCATTCGATCGTTCGCGAGCATCAGCCAC 107412
 QY 1141 TTGGTTCCTGCTTCTTTTGTGATCTGCACATCCTCGTGTGGCGCTTCCAGTGGAGTGTG 1200

ORIGIN

QY	541	CTACATCAGTAGGTGCGTGGCTGGATCGTACGATAACGAGGGGATTTGCCATATTCGCGCT	600
DB	130530	CTACATCAGTAGGTGCGTGGCTGGATCGTACGATAACGAGGGGATTTGCCATATTCGCGCT	130471
QY	601	GCAGTTCCACTACTTTCCTGTGGGTGCGCTCAGTGAAGACTGGATCCGTGTTCTGTGGTCGCG	660
DB	130470	GCAGTTCCACTACTTTCCTGTGGGTGCGCTCAGTGAAGACTGGATCCGTGTTCTGTGGTCGCG	130411
QY	661	CGGAGCGGCTTTGTGCTTACTTCTACATGGTGTCCGCTCGGGGTGGCTACGTGTTCAATCAT	720
DB	130410	CGGAGCGGCTTTGTGCTTACTTCTACATGGTGTCCGCTCGGGGTGGCTACGTGTTCAATCAT	130351
QY	721	CAACTGTATACCCCTGCACGCTCTTCGTACTGCTCATTTATGGGAGAGTACTCGCCGGCT	780
DB	130350	CAACTGTATACCCCTGCACGCTCTTCGTACTGCTCATTTATGGGAGAGTACTCGCCGGCT	130291
QY	781	GCTGACCAAGCTACAGACCTTCTACATCTGGGACTGCTGTTCTCCATGCAGATCCCTTT	840
DB	130290	GCTGACCAAGCTACAGACCTTCTACATCTGGGACTGCTGTTCTCCATGCAGATCCCTTT	130231
QY	841	CGTGGGATTCAAACCGATACGCACACAGTGAACATACGCTGCCTGGGAGTGTGTGCT	900
DB	130230	CGTGGGATTCAAACCGATACGCACACAGTGAACATACGCTGCCTGGGAGTGTGTGCT	130171
QY	901	CCTTATGGCCGTGGCCACTTTCGCGCAATTTGCAGTCCGTCTCTGTCGCGCAACGAGTTCCG	960
DB	130170	CCTTATGGCCGTGGCCACTTTCGCGCAATTTGCAGTCCGTCTCTGTCGCGCAACGAGTTCCG	130111
QY	961	GAAGCTGTTTCATCGTCGGCGGATTCGCTGGTGGCGGTGGCGCTTGTTCGCGCGTGTGCT	1020
DB	130110	GAAGCTGTTTCATCGTCGGCGGATTCGCTGGTGGCGGTGGCGCTTGTTCGCGCGTGTGCT	130051
QY	1021	GCTCACCATGCTGGCGTGTGTGGCCCCGTGGAGTGGACGCTTCTACTCGCTGTGGGATAC	1080
DB	130050	GCTCACCATGCTGGCGTGTGTGGCCCCGTGGAGTGGACGCTTCTACTCGCTGTGGGATAC	129991
QY	1081	TGGCTACGCCAAGATCCACATTCCCATCATTTGCATCCGTCGCGAGCATCAGCCACAC	1140
DB	129990	TGGCTACGCCAAGATCCACATTCCCATCATTTGCATCCGTCGCGAGCATCAGCCACAC	129931
QY	1141	TTGGTTCTCGTCTCTTCTTTGATCTGCACATCCTCGTGTGGCGCTTCCCATGGGAGTGTG	1200
DB	129930	TTGGTTCTCGTCTCTTCTTTGATCTGCACATCCTCGTGTGGCGCTTCCCATGGGAGTGTG	129871
QY	1201	GTACTGCATCAAGCAGATCAACGACGACGCGGCTTTTCGTGGTGCCTGTATGCCATCAGTGC	1260
DB	129870	GTACTGCATCAAGCAGATCAACGACGACGCGGCTTTTCGTGGTGCCTGTATGCCATCAGTGC	129811
QY	1261	GCTTTACTTCGCTGGTGTGATGGTTCATGTTACCCCTCACGCCGCTGGTGTGCAAT	1320
DB	129810	GCTTTACTTCGCTGGTGTGATGGTTCATGTTACCCCTCACGCCGCTGGTGTGCAAT	129751
QY	1321	GCTGCGCGGAGTGGCTTTTCGGGACTGTGTGGATGTGTTCTCTGCACAGAGATTCGTCMAA	1380
DB	129750	GCTGCGCGGAGTGGCTTTTCGGGACTGTGTGGATGTGTTCTCTGCACAGAGATTCGTCMAA	129691
QY	1381	GGGATGGSCACAGCCATTAAGCGGACGCCACCGAAGTGGATGAAGCTGAGGATTCATTCGA	1440
DB	129690	GGGATGGSCACAGCCATTAAGCGGACGCCACCGAAGTGGATGAAGCTGAGGATTCATTCGA	129631
QY	1441	GAAGAAGACGCTGTACGACAAGG	1463
DB	129630	GAAGAAGACGCTGTACGACAAGG	129608

RESULT 8

AC018145	75650 bp	DNA linear	HTG 09-DEC-1999
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.		
LOCUS	AC018145		
ACCESSION	AC018145		
VERSION	AC018145.1	GI:8553046	
KEYWORDS	HTG; HTGS PHASE2.		
SOURCE	Drosophila melanogaster (fruit fly)		

TITLE Detection kits, such as nucleic acid arrays, for detecting the
expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL Patent: WO 01/042-A 24553 27-SEP-2001;
PE Corporation (NY) (US)
FEATURES Location/Qualifiers
source 1..2785
/organism="Drosophila sp."
/mol_type="unassigned DNA"
/db_xref="taxon:7242"

ORIGIN

Query Match 10.3%; Score 249; DB 6; Length 2785;
Best Local Similarity 100.0%; Pred. No. 5.3e-136;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1876 ATCTTCACGCTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGCGGA 1935
Db |||||||
2785 ATCTTCACGCTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGCGGA 2726
QY 1936 TGATATCAACAAAGTTTCTGTGGATGCTCCGAATTGCTGAGGAGAGCATCCCAAGGACAT 1995
Db |||||||
2725 TGATATCAACAAAGTTTCTGTGGATGCTCCGAATTGCTGAGGAGAGCATCCCAAGGACAT 2666
QY 1996 TAAGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGATAGATCCCGAAGTGCTCC 2055
Db |||||||
2665 TAAGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGATAGATCCCGAAGTGCTCC 2606
QY 2056 GGCCTGCTCAACTGCTTATGTCACAAATTAGCTACTACAGATCCGGGAATTGAAGTT 2115
Db |||||||
2605 GGCCTGCTCAACTGCTTATGTCACAAATTAGCTACTACAGATCCGGGAATTGAAGTT 2546
QY 2116 GGACTACAG 2124
Db |||||||
2545 GGACTACAG 2537

RESULT 11
AF548377
LOCUS 297 bp DNA linear PLN 22-OCT-2002
DEFINITION Talaromyces emersonii oligosaccharyl transferase gene, partial cds.
ACCESSION AF548377
VERSION AF548377.1 GI:24210987
KEYWORDS
SOURCE Talaromyces emersonii
ORGANISM Talaromyces emersonii
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Talaromyces.
Crawley, K., Murray, P.G., Mc Loughlin, L.P. and Tuohy, M.G.
Cloning of oligosaccharyl transferase from Talaromyces emersonii
Unpublished
2 (bases 1 to 297)
AUTHORS Crawley, K., Murray, P.G., Mc Loughlin, L.P. and Tuohy, M.G.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-2002) Biochemistry, National University of
Ireland Galway, Galway, Ireland
FEATURES Location/Qualifiers
source 1..297
/organism="Talaromyces emersonii"
/mol_type="genomic DNA"
/db_xref="taxon:68825"
mRNA <1..-297
CDS <1..-297
/product="oligosaccharyl transferase"
/codon_start=1
/product="oligosaccharyl transferase"
/protein_id="AA52088.1"
/db_xref="GI:24210988"
/translation="GVIAPWGFRFYSIMDTGVAKIHPIIASVSEHPTAWPAFFDL
NFLVWLFPAQVYVCFDLKDEHFVIVIAVLSYFAGVWVRLMLTIPVVCVAA"

ORIGIN

Query Match 1.2%; Score 29; DB 8; Length 297;

Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1085 TAGCCCAAGATCCACATTCCTCCATTCATTC 1113
Db |||||||
52 TAGCCCAAGATCCACATTCCTCCATTCATTC 80

RESULT 12

AC118338
LOCUS 218635 bp DNA linear HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-167E4, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC118338
AC118338.3 GI:25138003
HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 218635)

REFERENCE

AUTHORS Muzny, D.Marie., Metzker, M.Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Y., Chen, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Louised, H., Lozada, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martin, R., McNeill, T.Z.,
McWhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidas, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwakoelam, O., Okwuonu, G., Olariunpago, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Shen, H.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shasman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, J., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmali, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von
Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished

TITLE
JOURNAL

REFERENCE	2 (bases 1 to 218635)	AC105647	263377 bp	DNA	linear	HTG 26-SEP-2002
AUTHORS	Worley,K.C.	AC105647				
TITLE	Direct Submission					
JOURNAL	Submitted (15-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
REFERENCE	3 (bases 1 to 218635)	AC105647.3	GI:23321658			
AUTHORS	Rat Genome Sequencing Consortium.	AC105647				
TITLE	Direct Submission					
JOURNAL	Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
COMMENT	On Nov 20, 2002 this sequence version replaced gi:23334685. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.					
	----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GWX Center clone name: CH230-167E4 ----- Summary Statistics Assembly program: Phrap; version 0.990329 Consensus quality: 207870 bases at least Q40 Consensus quality: 209962 bases at least Q30 Consensus quality: 211207 bases at least Q20 Estimated insert size: 213113; sum-of-contigs estimation Quality coverage: 6x in Q20 bases; sum-of-contigs estimation ----- * NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html). * NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * 1 215127: contig of 215127 bp in length * 215128 215227: gap of unknown length * 215228 216301: contig of 1074 bp in length * 216302 216401: gap of unknown length * 216402 218635: contig of 2234 bp in length. Location/Qualifiers 1. 218635 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-167E4"					
FEATURES	source					
ORIGIN	Query Match 1.28; Score 29; DB 2; Length 218635; Best Local Similarity 100.0%; Pred. No. 0.00023; Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	2389 GAATTAACCCAAAAAATAAAAAA 2417					
DB	83070 GAATTAACCCAAAAAATAAAAAA 83098					

TITLE
 JOURNAL
 COMMENT
 Direct Submission
 Submitted (26-SEP-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Sep 26, 2002 this sequence version replaced gi:21736544.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the
 sequence may extend beyond the ends of the clone and there may be
 contigs that consist entirely of whole genome shotgun sequence
 reads. Both end sequences and whole genome shotgun sequence only
 contigs will be indicated in the feature table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GNX
 Center clone name: CH230-99P13
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 197284 bases at least Q40
 Consensus quality: 203353 bases at least Q30
 Consensus quality: 207410 bases at least Q20
 Estimated insert size: 227312; sum-of-contigs estimation
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N. But the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 69640: contig of 69640 bp in length
 * 69641 69740: gap of unknown length
 * 69741 81842: contig of 12102 bp in length
 * 81843 81942: gap of unknown length
 * 81943 86675: contig of 4733 bp in length
 * 86676 86775: gap of unknown length
 * 86776 97506: contig of 10731 bp in length
 * 97507 97606: gap of unknown length
 * 97607 136072: contig of 38466 bp in length
 * 136073 136172: gap of unknown length
 * 136173 257131: contig of 120959 bp in length
 * 257132 257231: gap of unknown length
 * 257232 259804: contig of 2573 bp in length
 * 259805 259905: gap of unknown length
 * 259906 263377: contig of 3473 bp in length.
 * 263377 Location/Qualifiers
 source
 1..263377
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-99P13"
 86776..88251
 /note="wgs_contig"
 97607..99118
 /note="wgs_contig"
 143266..144894
 /note="wgs_contig"
 180092..181612
 /note="wgs_contig"
 220230..221431
 /note="wgs_contig"
 245895..249101
 /note="wgs_contig"
 ORIGIN
 Query Match 1.2%: Score 29; DB 2; Length 263377;
 Best Local Similarity 100.0%; Pred.No.0.00023;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2389 GAATATACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
 Db 197394 GAATATACCCCAAAAAAAAAAAAAAAAAAAAAA 197422
 RESULT 14
 AC116217 285045 bp DNA linear HTG 10-OCT-2002
 LOCUS Rattus norvegicus clone CH230-111110, *** SEQUENCING IN PROGRESS
 DEFINITION *** 7 unordered pieces.
 AC116217
 AC116217.3 GI:23664575
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 285045)
 Muzny, D.Marie., Metzker, M.Lee., Abramzon, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Chen, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Detamo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gregeorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kratt, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensu, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Minet, G., Minje, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Naif, D.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwackelme, O., Okwuonu, G., Olarinpo, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajd, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wlezyk, R., Woodden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.
 Direct Submission

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 285045)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 285045)
AUTHORS Direct Submission
TITLE Direct Submission
JOURNAL Submitted (10-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Oct 10, 2002 this sequence version related gi:21671715.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center Project name: GNLF
Center Clone name: CH230-11110
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 240039 bases at least Q40
Consensus quality: 244142 bases at least Q30
Consensus quality: 246617 bases at least Q20
Estimated insert size: 248451; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 129760: contig of 129760 bp in length
* 129761 129860: gap of unknown length
* 129861 129868: contig of 13848 bp in length
* 268509 268608: gap of unknown length
* 268609 278238: contig of 9630 bp in length
* 278239 278338: gap of unknown length
* 278339 279469: contig of 1131 bp in length
* 279470 279569: gap of unknown length
* 279570 281290: contig of 1721 bp in length
* 281291 281390: gap of unknown length
* 281391 282813: contig of 1423 bp in length
* 282814 282913: gap of unknown length
* 282914 285045: contig of 2132 bp in length.
* Location/Qualifiers
* 1..285045
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clones="CH230-11110"
* 1..1031
* /note="wgs_end_extension"
*
misc_feature
source
FEATURES
source
1..1968
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEP21211"
/cell_line="HepG2"

```

```

clone_end:Sp6"
1082..2099
/note="wgs_end_extension"
clone_end:Sp6"
4359..5234
/note="clone_boundary"
clone_end:Sp6"
site:EcoRI
end_sequence:BH339077"
4442..119103
/note="clone_boundary"
clone_end:Sp6"
site:EcoRI
end_sequence:BH339077"
59773..76032
/note="clone_boundary"
clone_end:T7
site:EcoRI
end_sequence:BH339051"
268609..270963
/note="wgs_end_extension"
clone_end:Sp6"
ORIGIN
Query Match 1.2%; Score 29; DB 2; Length 285045;
Best Local Similarity 100.0%; Pred.No. 0.00023;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2389 GAAATTAACCCAAAAA 2417
Db 207074 GAAATTAACCCAAAAA 207102
RESULT 15
AK074363 1968 bp mRNA linear PRI 12-SEP-2003
LOCUS AK074363
DEFINITION Homo sapiens cDNA FLJ23783 fis, clone HEP21211, highly similar to
Homo sapiens tripartite motif protein TRIM5 isoform alpha mRNA.
ACCESSION AK074363
VERSION AK074363.1 GI:18576945
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
TITLE Unpublished
JOURNAL NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan, cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction:
5'- & 3'-end one pass sequencing: Department of Virology and Human
Genome Center, Institute of Medical Science, University of Tokyo
(partly supported by Science and Technology Agency).
FEATURES
source
1..1968
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEP21211"
/cell_line="HepG2"

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 00:53:33 ; Search time 1624 Seconds
(without alignments)
7812.721 Million cell updates/sec

Title: US-10-028-384-7

Perfect score: 2417

Sequence: 1 tctagcgaagtgtgtcg.....ccaaaaaaaaaaaaaaaaa 2417

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 4134886 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq 23Sep04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2417	100.0	2417	10	ADD94789	ADD94789 Drosophil
2	2409	99.7	2699	4	ABL13247	ABL13247 Drosophil
3	1463	60.5	4922	4	ABL13246	ABL13246 Drosophil
4	696	28.8	2953	4	ABL18224	ABL18224 Drosophil
5	249	10.3	2785	4	ABL18208	ABL18208 Drosophil
6	28	1.2	442	6	ABQ59909	ABQ59909 Human col
7	27	1.1	391	4	AAL26051	AAL26051 Human bre
8	27	1.1	401	4	AAL08306	AAL08306 Human bre
9	27	1.1	471	4	AAL16849	AAL16849 Human bre
10	26	1.1	2855	4	ABL02795	ABL02795 Drosophil
11	26	1.1	3248	10	ADC37655	ADC37655 Human ang
12	26	1.1	6153	4	ABL02794	ABL02794 Drosophil
13	25	1.0	253	5	ABV61073	ABV61073 Human pro
14	25	1.0	282	5	ABV61368	ABV61368 Human pro
15	25	1.0	340	5	ABV43386	ABV43386 Human pro
16	25	1.0	340	5	ABV34528	ABV34528 Human pro
17	25	1.0	354	6	ABZ08446	ABZ08446 Human leu
18	25	1.0	429	5	ABV13410	ABV13410 Human pro
19	25	1.0	434	12	ADN13434	ADN13434 Human pro
20	25	1.0	513	5	ADI75635	ADI75635 Human ova
21	25	1.0	513	5	ADI69289	ADI69289 Human ova

22	25	1.0	575	4	RAF33252	RAF33252 Human sec
23	25	1.0	583	5	ADL40871	ADL40871 Human ova
24	25	1.0	605	5	ABV58657	ABV58657 Human pro
25	25	1.0	659	6	ABQ28115	ABQ28115 Oligonuc
26	25	1.0	659	6	ABQ28114	ABQ28114 Oligonuc
27	25	1.0	867	10	ADG32855	ADG32855 Human DNA
28	25	1.0	1023	6	ABT09673	ABT09673 Human PAL
29	25	1.0	1033	6	ABT09672	ABT09672 Human PAL
30	25	1.0	1033	6	ABT09672	ABT09672 Haematopo
31	25	1.0	8044	4	AS46625	AS46625 Tunour su
32	25	1.0	8244	10	ADB54282	ADB54282 Fretreat
33	24	1.0	130	5	ABV61447	ABV61447 Human pro
34	24	1.0	142	4	PAL26330	PAL26330 Human bre
35	24	1.0	143	5	ADL44859	ADL44859 Human ova
36	24	1.0	155	4	ADL19280	ADL19280 Human bre
37	24	1.0	167	5	ADI76056	ADI76056 Human ova
38	24	1.0	167	5	ADI69719	ADI69719 Human ova
39	24	1.0	175	8	ABX51070	ABX51070 Bovine ES
40	24	1.0	187	4	AAL08585	AAL08585 Human bre
41	24	1.0	200	6	ABQ56626	ABQ56626 Human col
42	24	1.0	211	5	ABV19351	ABV19351 Human pro
43	24	1.0	216	5	ABV61105	ABV61105 Human pro
44	24	1.0	222	5	ABV61623	ABV61623 Human pro
45	24	1.0	224	5	AS29077	AS29077 cDNA enco

ALIGNMENTS

RESULT 1

ADD94789

ID ADD94789 standard; DNA; 2417 BP.

AC ADD94789;

XX 29-JAN-2004 (first entry)

XX Drosophila melanogaster STT3 gene sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;

XX major histocompatibility complex; human leukocyte antigen; HLA;

XX cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;

XX lung cancer; intestine cancer; sarcoma; prostate cancer;

XX testicular cancer; breast cancer; melanomas; pancreatic cancer;

XX haematological cancer; immune response; lymphoid cell proliferation;

XX autoimmune disease; transplant rejection; SIMP-derived peptide;

XX fruit fly; gene; ds; STT3.

XX Drosophila melanogaster.

XX WC2003054008-A2.

XX 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

XX 20-DEC-2001; 2001US-00028384.

XX (COMP-) COMPATIGENE INC.

XX Perreault C, McBride K;

XX WPI; 2003-559122/52.

XX P-PSDB; ADD94790.

XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic

XX acids and proteins, useful for diagnosing and treating cancers, e.g. lung

XX or breast cancer, or for suppressing an immune response in an autoimmune

XX disease.

XX Claim 6; SEQ ID NO 7; 66pp; English.

XX This invention relates to a novel isolated or purified human protein,

XX

termed source of immunodominant major histocompatibility complex (MHC) - associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytotoxic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestinal cancer, sarcomas, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the *Protophila melanogaster* SMT3 gene which is related to the invention. Note: this sequence does not appear in the CC specification but was obtained by the indexer from GenBank.

XX Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;
SQ

Query Match 100.0%; Score 2417; DB 10; Length 2417;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAAGCGAAGATGTGTCGTTCGATTTCAGATCGGTTAATTTTCGAGTTACTGGCTG 60
Db 1 TCTAAGCGAAGATGTGTCGTTCGATTTCAGATCGGTTAATTTTCGAGTTACTGGCTG 60

QY 61 GAAATTGGGACATGAATTCGGACGCCGAAGATGCTGAACAGCAGAGTGGCTGCACAG 120
Db 61 GAAATTGGGACATGAATTCGGACGCCGAAGATGCTGAACAGCAGAGTGGCTGCACAG 120

QY 121 CCTATATCACTTCGCCATCTGCTAATCGGCTGGCTGGCGGAGTTTCCTCTCGCTCTT 180
Db 121 CTTAATCACTTCGCCATCTGCTAATCGGCTGGCTGGCGGAGTTTCCTCTCGCTCTT 180

QY 181 CGCGCTCATCCGTTTCGAGTCGATTATCCATGAGTTTGATCCGTGTTCAACTACCGGCG 240
Db 181 CGCGCTCATCCGTTTCGAGTCGATTATCCATGAGTTTGATCCGTGTTCAACTACCGGCG 240

QY 241 CACCGCTACATGGTGCGAGATGGTGGTCAACTTCCTCACTGGTTCGACGAGCGGC 300
Db 241 CACCGCTACATGGTGCGAGATGGTGGTCAACTTCCTCACTGGTTCGACGAGCGGC 300

QY 301 ATGGTATCCGCTCGCAGGATTTGGCGGTAACCGTCTATCCCGGCTCATGATTACGTC 360
Db 301 ATGGTATCCGCTCGCAGGATTTGGCGGTAACCGTCTATCCCGGCTCATGATTACGTC 360

QY 361 CGCGCGAATCCATTGGCTGTCGACGTACTCACATACCGTCCATATTCGTGACATCTG 420
Db 361 CGCGCGAATCCATTGGCTGTCGACGTACTCACATACCGTCCATATTCGTGACATCTG 420

QY 421 CGTGTTCCTGGCGCGAATTCCTAGTGGCTGACCTCCATCTCACTACTCTCTGACCAA 480
Db 421 CGTGTTCCTGGCGCGAATTCCTAGTGGCTGACCTCCATCTCACTACTCTCTGACCAA 480

QY 481 GGAGCTGTGGTCCGCGGCGCGGACCTTCCTGGCGCGCAGCTTCATCGCCATCGTGCCTGG 540
Db 481 GGAGCTGTGGTCCGCGGCGCGGACCTTCCTGGCGCGCAGCTTCATCGCCATCGTGCCTGG 540

QY 541 CTACATCATGATAGTGGCTGGATCGTACGATAACGAGGGCAATTCGCCCTT 600
Db 541 CTACATCATGATAGTGGCTGGATCGTACGATAACGAGGGCAATTCGCCCTT 600

QY 601 GCAGTTCACTTCTCTGTGGTGGCTCAGTGAAGACTGGATCCGTGTCTGTGTGCGC 660
Db 601 GCAGTTCACTTCTCTGTGGTGGCTCAGTGAAGACTGGATCCGTGTCTGTGTGCGC 660

QY 661 CGCAGCGGTTTGCTACTTCTACATGTTGTCGCTGGGTTGGCTACGTGTTCAATCAT 720
Db 661 CGCAGCGGTTTGCTACTTCTACATGTTGTCGCTGGGTTGGCTACGTGTTCAATCAT 720

QY 721 CAACTTGATACCCCTGCACGCTTTCGTACTGCTCATTTATGGGCAGAGTACTCGCCGCTCT 780

Db 1801 GTGGAACAATAGTCATAGACGCTGGTTGGCAAGCAATGCTTCAACCGAGGAGATC 1860
QY 1861 CTACGAATATTATGACATCTTTGACGTGACCTACGTTTGGTGATCTTTGGCGGTGAT 1920
Db 1861 CTACGAATATTATGACATCTTTGACGTGACCTACGTTTGGTGATCTTTGGCGGTGAT 1920
QY 1921 CGGCTATTCTGCGGATGATATCAACAAGTTCCTGTGGATGGTCCGAATTGCTGAGGAGA 1980
Db 1921 CGGCTATTCTGCGGATGATATCAACAAGTTCCTGTGGATGGTCCGAATTGCTGAGGAGA 1980
QY 1981 GCATCCCAAGGACATTAAGGAAGGATTAATTTACCGACCGCGGTGAATTCAGGGTAGA 2040
Db 1981 GCATCCCAAGGACATTAAGGAAGGATTAATTTACCGACCGCGGTGAATTCAGGGTAGA 2040
QY 2041 TCCGAAAGGTGCTCCGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATT 2100
Db 2041 TCCGAAAGGTGCTCCGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATT 2100
QY 2101 CGGGAATTAAGTTGGACTACAGAGTCCATCTGGATATGATCGCACAGTAACGCCGT 2160
Db 2101 CGGGAATTAAGTTGGACTACAGAGTCCATCTGGATATGATCGCACAGTAACGCCGT 2160
QY 2161 CATTGGGAATTAAGGACTTCGATCTGACCTACCTGGAGAGGCTTACACCAAGCACTG 2220
Db 2161 CATTGGGAATTAAGGACTTCGATCTGACCTACCTGGAGAGGCTTACACCAAGCACTG 2220
QY 2221 GCTTGTTCGATCTATAGGTTGAAGAGCCGATGATTCATAGACCATCACTGAAGAC 2280
Db 2221 GCTTGTTCGATCTATAGGTTGAAGAGCCGATGATTCATAGACCATCACTGAAGAC 2280
QY 2281 CAAGGAGAGAACGATTCTCCAGCAAACTTCATTTCCGAGAAAGACTCTTAAGCGTGC 2340
Db 2281 CAAGGAGAGAACGATTCTCCAGCAAACTTCATTTCCGAGAAAGACTCTTAAGCGTGC 2340
QY 2341 GGGCTACATAGCAAAACCGCGTGTGTTTAAAGGAAACGAACTTGAATAAACCCCA 2400
Db 2341 GGGCTACATAGCAAAACCGCGTGTGTTTAAAGGAAACGAACTTGAATAAACCCCA 2400
QY 2401 AAAAAAAAAAAAAA 2417
Db 2401 AAAAAAAAAAAAAA 2417

RESULT 2

ABL13247

ID ABL13247 standard; cDNA; 2699 BP.

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX AC ABL13247;

XX

PT

PR

XX

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

SQ

New isolated nucleic acid detection reagent for detecting 1000 or more genes from *Drosophila* and for elucidating cell signalling and cell-cell interactions.

Claim 1; SEQ ID NO 34223; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO.int/pub/pub/published_pct_sequences

Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other;

Query Match 99.7%; Score 2409; DB 4; Length 2699;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAAGCGAAGATGTGCTTCATTCAGATCGGTTATAATTTTCAGATTACTGGCTG 60

Db 10 TCTAAGCGAAGATGTGCTTCATTCAGATCGGTTATAATTTTCAGATTACTGGCTG 69

QY 61 GAATGGGACATGATCGACCCGAGAGTGTGAACAGAGGTGGTGGTACAGCAG 120

Db 70 GAATGGGACATGATCGACCCGAGAGTGTGAACAGAGGTGGTGGTACAGCAG 129

QY 121 CTAATACCTTCGGCATCTCTAATCGCTGGCTGGCGGATTTCTCTCGGCTCTT 180

Db 130 CTAATACCTTCGGCATCTCTAATCGCTGGCTGGCGGATTTCTCTCGGCTCTT 189

QY 181 CCGCTCATCGCTTCGAGTGCATTCATCGAGTTTTCATCGTGGTTCAATACCGGC 240

Db 190 CCGCTCATCGCTTCGAGTGCATTCATCGAGTTTTCATCGTGGTTCAATACCGGC 249

QY 241 CACCGCTTACATGTCGAGATGTTGGCGGTACCGTCTATCCCGGCTGTGATTACGTC 300

Db 250 CACCGCTTACATGTCGAGATGTTGGCGGTACCGTCTATCCCGGCTGTGATTACGTC 309

QY 301 ATGGTATCCGCTCGCAGGATTTGGCGGTACCGTCTATCCCGGCTGTGATTACGTC 360

Db 310 ATGGTATCCGCTCGCAGGATTTGGCGGTACCGTCTATCCCGGCTGTGATTACGTC 369

QY 361 CCGCGGAATCCATTGGCTGCTGCAGTACTCAACATACCGGTCCATTCGTGACATCTG 420

Db 370 CCGCGGAATCCATTGGCTGCTGCAGTACTCAACATACCGGTCCATTCGTGACATCTG 429

QY 421 CGTGTTCCTGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGTCGACCA 480

Db 430 CGTGTTCCTGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGTCGACCA 489

QY 481 GGAGCTGTGTCGCGGGCGCGGCTCTTCGCGCGAGTTCATCCCATTCGTCCTGG 540

Db 490 GGAGCTGTGTCGCGGGCGCGGCTCTTCGCGCGAGTTCATCCCATTCGTCCTGG 549

QY 541 CTACATCAGTAGGTGCTGGTGGTATCGTAGTAACGAGGCAATTCGCATATTCGCGCT 600

Db 550 CTACATCAGTAGGTGCTGGTGGTATCGTAGTAACGAGGCAATTCGCATATTCGCGCT 609

QY 601 GCAGTTCACCTTCTGTTGGTGGCTCAGTGAAGACTGGATTCGTTTCTGGTGGCG 660

Db 610 GCAGTTCACCTTCTGTTGGTGGCTCAGTGAAGACTGGATTCGTTTCTGGTGGCG 669

QY 661 GCGAGCGCTTGTCTTACTTCTACATGTTGTCGCGCTGGGTCGCTACGTGTTTCATCAT 720

Db 670 GCGAGCGCTTGTCTTACTTCTACATGTTGTCGCGCTGGGTCGCTACGTGTTTCATCAT 729

QY 721 CAACCTGATACCCCTGCACTCTTCGTACTGCTCATTTATGGGAGGTAATTCGCGCGCTCT 780

Db	730	CAACCTGATACCCCTGACGCTTCCTGCTACTGCTCATTTATGGCAGGTACTCGCCGCGTCT	789
Qy	781	GCTGACCAAGCTACAGCAACCTTCTACATCTCTGGGACTGCTGTTCTCATACGACAGATCCCGCTT	840
Db	790	GCTGACCAAGCTACAGCAACCTTCTACATCTCTGGGACTGCTGTTCTCATGACAGATCCCGCTT	849
Qy	841	CGTGGGATTCCAAACCGATACGACCAAGTGAACATGCGCTGCGCTGGAGTGTGTTGTGCT	900
Db	850	CGTGGGATTCCAAACCGATACGACCAAGTGAACATGCGCTGCGCTGGAGTGTGTTGTGCT	909
Qy	901	CGTTATGCCCCTGGCCACCTTGGCCCATTTGCAAGTCCGTCTGTCTCGCCAAACGAGTTCGG	960
Db	910	CGTTATGCCCCTGGCCACCTTGGCCCATTTGCAAGTCCGTCTGTCTCGCCAAACGAGTTCGG	969
Qy	961	GAAGCTGTTCATCGTCCGCGGATGCTGTGTGGCGCTTGGCGTCTTTGTGGCGCTGTGGT	1020
Db	970	GAAGCTGTTCATCGTCCGCGGATGCTGTGTGGCGCTTGGCGTCTTTGTGGCGCTGTGGT	1029
Qy	1021	GCTCACCAATGCTGGGCGTGTGGCCCGCTGGAGTGAACGCTTCTACTCGCTGTGGGATAC	1080
Db	1030	GCTCACCAATGCTGGGCGTGTGGCCCGCTGGAGTGAACGCTTCTACTCGCTGTGGGATAC	1089
Qy	1081	TGGCTACGCCAGATCCACATTCCTCATCTGTCATCCGCTCGGAGCATCAGCCACCAC	1140
Db	1090	TGGCTACGCCAGATCCACATTCCTCATCTGTCATCCGCTCGGAGCATCAGCCACCAC	1149
Qy	1141	TTGGTTCCTGTTCTTCTTGATCTGCACATCTGTGTGCGCTTCCCAAGTGGGAGTGTG	1200
Db	1150	TTGGTTCCTGTTCTTCTTGATCTGCACATCTGTGTGCGCTTCCCAAGTGGGAGTGTG	1209
Qy	1201	GTACTGTCATCAAGCAGATCAACGACGAGCGCTTTTCGTGTGCTGTACGCCATCAGTGC	1260
Db	1210	GTACTGTCATCAAGCAGATCAACGACGAGCGCTTTTCGTGTGCTGTGTAGCCATCAGTGC	1269
Qy	1261	GGTTTACTTCCCTCGTGTGATGCGGTGTGATGTGACCTTCACGCCGCTGGTGTGCAT	1320
Db	1270	GGTTTACTTCCCTCGTGTGATGCGGTGTGATGTGACCTTCACGCCGCTGGTGTGCAT	1329
Qy	1321	GCTGCCCGAGTGGCCCTTTTCGGGACTGTTTGGAGTGTCTCTGCAAGAGGATTCGTTCAA	1380
Db	1330	GCTGCCCGAGTGGCCCTTTTCGGGACTGTTTGGAGTGTCTCTGCAAGAGGATTCGTTCAA	1389
Qy	1381	CGGAATTGGGCAACGCAATAAGCGACGACCCGAAGTGGANGAAGCTGAGGATTCATTTGA	1440
Db	1390	CGGAATTGGGCAACGCAATAAGCGACGACCCGAAGTGGANGAAGCTGAGGATTCATTTGA	1449
Qy	1441	GAAGAAGACGCTGTACACAAAGCTGCGAAGCTGAAGCATCGTACTAAGACATGATGCCCA	1500
Db	1450	GAAGAAGACGCTGTACACAAAGCTGCGAAGCTGAAGCATCGTACTAAGACATGATGCCCA	1509
Qy	1501	CGAGGATCTGGCTGAGCTCCAACTGAAGAGTATGTATTTTGGCCGCTTTAAATGCT	1560
Db	1510	CGAGGATCTGGCTGAGCTCCAACTGAAGAGTATGTATTTTGGCCGCTTTAAATGCT	1569
Qy	1561	GTGTATGATGTTCCGCTGTCCACTGCACTGGGTGACCGCAATGCTACTCCAGTCCCTC	1620
Db	1570	GTGTATGATGTTCCGCTGTCCACTGCACTGGGTGACCGCAATGCTACTCCAGTCCCTC	1629
Qy	1621	CATTGTCTTGGCTTCCAAACAGTCAAGTGGATCCCGCAACATTTTAGACGATTTTCAG	1680
Db	1630	CATTGTCTTGGCTTCCAAACAGTCAAGTGGATCCCGCAACATTTTAGACGATTTTCAG	1689
Qy	1681	AGAGGCTTACTCTGSGGCTTTTCGCAAGCACTGCGCATGATGCTCGCGTTATGTCTTGGTG	1740
Db	1690	AGAGGCTTACTCTGSGGCTTTTCGCAAGCACTGCGCATGATGCTCGCGTTATGTCTTGGTG	1749
Qy	1741	GGATTACGGATACAGATACGGGATTCGCAACAGACGAGCTAGTGGATATATATAC	1800
Db	1750	GGATTACGGATACAGATACGGGATTCGCAACAGACGAGCTAGTGGATATATATATAC	1809
Qy	1801	GTGGAAACAATAGTCACATAGCGCTGGTTGGCAAGGCAATGCTTCCAAACGAGGAGAGTC	1860

Db	1810	GTGGAAACAATAGT	CACATAGCGCTGTGTGGCAAGCAATGTCTTCAACCGAGGAGAAATC	1869
Qy	1861	CTACGAAATATGACATCTCTT	GACGTGGACACTCGTTTGGTGTGATCTTTTGGCGGTGTGAT	1920
Db	1870	CTACGAAATATGACATCTCTT	GACGTGGACACTCGTTTGGTGTGATCTTTTGGCGGTGTGAT	1929
Qy	1921	CGGCTATTCTGGCGATGATAT	CAACAAGTTCTCTGTGGATGGTCCGAAATTCCTGAGGGAGA	1980
Db	1930	CGGCTATTCTGGCGATGATAT	CAACAAGTTCTCTGTGGATGGTCCGAAATTCCTGAGGGAGA	1989
Qy	1981	GCATCCCAAGGACATTAAAGAA	AGCGGATTACTTTTACCAGCCGCGTGAATTCAGGGTAGA	2040
Db	1990	GCATCCCAAGGACATTAAAGAA	AGCGGATTACTTTTACCAGCCGCGTGAATTCAGGGTAGA	2049
Qy	2041	TGCGAAGGTGCTCCGCGCTGCT	CAACTGCCTTATGTACAAATTAAGCTACTACAGATT	2100
Db	2050	TGCGAAGGTGCTCCGCGCTGCT	CAACTGCCTTATGTACAAATTAAGCTACTACAGATT	2109
Qy	2101	CGGGAAATTGAAGTTGGACTAC	AGAGGTCCTCTCGATATGATCCGACACGTAAACGCCGT	2160
Db	2110	CGGGAAATTGAAGTTGGACTAC	AGAGGTCCTCTCGATATGATCCGACACGTAAACGCCGT	2169
Qy	2161	CATTGGGAATAAGGACTTCGAT	CTGTGACCTACCTGGAGGAGGCTTACACCAAGAACACTG	2220
Db	2170	CATTGGGAATAAGGACTTCGAT	CTGTGACCTACCTGGAGGAGGCTTACACCAAGAACACTG	2229
Qy	2221	GTTTCTTCGCATCTPATAGGTG	AAAGCGCGATGAGTTCAATAGACCATCACTGAAGAC	2280
Db	2230	GTTTCTTCGCATCTPATAGGTG	AAAGCGCGATGAGTTCAATAGACCATCACTGAAGAC	2289
Qy	2281	CAAGGAGAGAACGATTCTCCT	CCACCAACTTCATTTTCGAGAAAGAACTCTTAAGCGTCGCAA	2340
Db	2290	CAAGGAGAGAACGATTCTCCT	CCACCAACTTCATTTTCGAGAAAGAACTCTTAAGCGTCGCAA	2349
Qy	2341	GGGCTACATACGAAACCGAC	CCGGTTCTGTTTAAAGGAAAAAGAAACCTTGAATAAACC	2400
Db	2350	GGGCTACATACGAAACCGAC	CCGGTTCTGTTTAAAGGAAAAAGAAACCTTGAATAAACC	2409
Qy	2401	AAAAA	AAAAA 2409	
Db	2410	AAAAA	AAAAA 2418	

RESULT 3

AB113246/c

ID ABL3246 standard; cDNA; 4922 BP.

RESULT 3
ABL13246/c
ID ABL13246 standard; cDNA: 4922 BP.

AX
AC ABL13246;

XX
DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster ex

XX

pharmaceutical; gene; ss.

OS *Drosophila melanogaster*.

XX
PN WO200171042-A2.

XX
PD
27-SEP-2001.

XX
DE
23-MAR-2001.

XX
DE

PR 11-JUL-2000; 2000US-00614150.

PA (PEKE) PE CORP NY.

XX
PI
Venter JC, Adams M

XX
DR WPI: 2001-656860/75-

DR P-PSDB; ABB69143.


```

PR 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
PA (PEKE ) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 6145; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
CC ABBS72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2953 BP; 924 A; 628 C; 648 G; 753 T; 0 U; 0 Other;
XX
Query Match 28.8%; Score 696; DB 4; Length 2953;
Best Local Similarity 100.0%; Pred. No. 1e-255;
Matches 696; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAAGCGAAGATGTGCTGCTGATTCAGATCGGTATATATTCGAGTTACTGGCTG 60
DB 696 TCTAAGCGAAGATGTGCTGCTGATTCAGATCGGTATATATTCGAGTTACTGGCTG 637
QY 61 GAATTTGGGACATGAATCGGACCGAAGATGCTGAACAGCAGGCTGGCTGACAGCAG 120
DB 636 GAATTTGGGACATGAATCGGACCGAAGATGCTGAACAGCAGGCTGGCTGACAGCAG 577
QY 121 CCTAATCACTTCGCAATCCTGTAATCGCTGGCTGGCGGATTTCTCTCGGCTTT 180
DB 576 CCTAATCACTTCGCAATCCTGTAATCGCTGGCTGGCGGATTTCTCTCGGCTTT 517
QY 181 CGCGCTCATCGCTTCGAGTCGATTAATCCATGATTTGATCCGCTGTTCAACTACCGGC 240
DB 516 CGCGCTCATCGCTTCGAGTCGATTAATCCATGATTTGATCCGCTGTTCAACTACCGGC 457
QY 241 CACGCGCTACATGTGTCAGATGTTGTTACAACTTCCTCACTGTTGTCAGAGCGGC 300
DB 456 CACGCGCTACATGTGTCAGATGTTGTTACAACTTCCTCACTGTTGTCAGAGCGGC 397
QY 301 ATGATATCCGCTCGCGAGGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTACGTC 360
DB 396 ATGATATCCGCTCGCGAGGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTACGTC 337
QY 361 CGCGGATATCAATTTGGCTGTCGACGTAATCAACATCCGCTCATATTCGTGATCTG 420
DB 336 CGCGGATATCAATTTGGCTGTCGACGTAATCAACATCCGCTCATATTCGTGATCTG 277
QY 421 CGTGTTCCTGGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480
DB 276 CGTGTTCCTGGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 217
QY 481 GGAGCTGTGTCGCGGCGCGCTCTTCGCGCGAGCTTCATCCCATGTCGCTCG 540
DB 216 GGAGCTGTGTCGCGGCGCGCTCTTCGCGCGAGCTTCATCCCATGTCGCTCG 157
QY 541 CTACATCAGTAGGTGGCTGGATCTGATCAACAGGAGGATTCGCTATTCGCTTC 600
DB 156 CTACATCAGTAGGTGGCTGGATCTGATCAACAGGAGGATTCGCTATTCGCTTC 97
QY 601 GCAGTTACCTTACTTCTCTGCTGGTGCCTCAGTGAAGACTGGATCCGCTTCTGCTCGC 660

```

```

DB 96 GCAGTTACCTACTTCTCTGCTGGTCCGCTCAGTGAAGACTGGATCCGCTTCTGCTCGC 37
QY 661 CGCAGCCGCTTTGCTCTACTTCTACATGCTGTCGC 696
DB 36 CGCAGCCGCTTTGCTCTACTTCTACATGCTGTCGC 1
RESULT 5
ABL18208/c
ID ABL18208 standard; DNA; 2785 BP.
XX
XX ABL18208;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 6097.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US009231.
PF
XX
XX 23-MAR-2000; 2000US-0191637P.
PR
XX
XX 11-JUL-2000; 2000US-00614150.
PR
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
PT
XX
XX Claim 1; SEQ ID NO 6097; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
CC ABBS72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2785 BP; 795 A; 644 C; 604 G; 742 T; 0 U; 0 Other;
XX
Query Match 10.3%; Score 249; DB 4; Length 2785;
Best Local Similarity 100.0%; Pred. No. 3.2e-85;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1876 ATCTCTTGACGTGAGTACGTTTGGTGATCTTTGGCGGTGTCGCTATTCGCGCA 1935
DB 2785 ATCTCTTGACGTGAGTACGTTTGGTGATCTTTGGCGGTGTCGCTATTCGCGCA 2726
QY 1936 TGATATCAACAGTTCTCTGTCGATGTCGATTCCTCAGGAGAGATCCCAAGGACAT 1995
DB 2725 TGATATCAACAGTTCTCTGTCGATGTCGATTCCTCAGGAGAGATCCCAAGGACAT 2666
QY 1996 TAAGGAAAGCGATTACTTTTACCGACCGCGGTGAATTCAGGGTAGATGCCAAGTGTCC 2055
DB 2665 TAAGGAAAGCGATTACTTTTACCGACCGCGGTGAATTCAGGGTAGATGCCAAGTGTCC 2606

```

QY 2056 GGCCTGCTCACTCCCTATGTACAAATTAAGCTACTACAGATTGCGGAATTGAAGTT 2115
Db 2605 GGCCTGCTCACTCCCTATGTACAAATTAAGCTACTACAGATTGCGGAATTGAAGTT 2546
QY 2116 GGCCTGCTCACTCCCTATGTACAAATTAAGCTACTACAGATTGCGGAATTGAAGTT 2546
Db 2545 GGCCTGCTCACTCCCTATGTACAAATTAAGCTACTACAGATTGCGGAATTGAAGTT 2546

RESULT 6
ABQ59909/c
ID ABQ59909 standard; cDNA; 442 BP.
XX
AC ABQ59909;
XX
XX
DT 02-AUG-2002 (first entry)
XX
DE Human colon cancer related nucleotide sequence SEQ ID NO:3604.
XX
KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
XX genetic analysis; diagnostic; antisense therapy; gene; ss.
XX Homo sapiens.
XX
PN WO200229086-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US030732.
XX
PR 02-OCT-2000; 2000US-0237271P.
XX
PA (FARB) BAYER CORP.
XX
PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI Thiagalingam A, Lewis MB;
XX
DR WPI; 2002-426115/45.
XX

XX New isolated nucleic acid that is differentially expressed in cancer
PT tissues useful for determining the presence of colon cancer in a cell or
PT tissue type, and in antisense therapy.
XX

PS Claim 1; Fig 1; 796pp; English.

CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC used in antisense therapy. An antibody immunoreactive with a polypeptide
CC encoded by (I) is useful for detecting cancer in a patient sample, and
CC for detecting the presence or absence of a polynucleotide encoded by a
CC nucleic acid which hybridizes to (I) in a cell. A probe/primer derived
CC from (I) can be used for determining the presence of a nucleic acid which
CC hybridizes to (I), and for determining the phenotype of cells in a sample
CC of cells from a patient. (I) is useful for determining the presence of
CC colon cancer in a cell or tissue type, for determining the presence or
CC state of other type of cancer, in antisense therapy, to generate
CC macroarrays on a solid surface, to identify a chromosome on which the
CC corresponding gene resides, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (I) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists
XX

SQ Sequence 442 BP; 102 A; 102 C; 91 G; 132 T; 0 U; 15 Other;

Query Match 1.2%; Score 28; DB 6; Length 442;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2390 AATTAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 34 AATTAACCCCAAAAAAAAAAAAAAAAAAAAAA 7

RESULT 7
AAL26051/c
ID AAL26051 standard; cDNA; 391 BP.
XX
AC AAL26051;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 18508.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
XX
PR 14-MAR-2000; 2000US-0189167P.
XX
PR 24-MAR-2000; 2000US-0192099P.
XX
PR 29-MAR-2000; 2000US-0193480P.
XX
PR 09-JUN-2000; 2000US-0205230P.
XX
PR 25-JUL-2000; 2000US-0211315P.
XX
PR 25-JUL-2000; 2000US-0220534P.
XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

PI WPI; 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer.

PS Claim 1; Page 3414; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity

XX Sequence 391 BP; 96 A; 65 C; 70 G; 160 T; 0 U; 0 Other;

Query Match 1.1%; Score 27; DB 4; Length 391;

Best Local Similarity 100.0%; Pred. No. 2.1;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2390 AATTAACCCCAAAAAAAAAAAAAAAAAAAAAA 2416

Db 29 AATTAACCCCAAAAAAAAAAAAAAAAAAAAAA 3

RESULT 8

AAL08306/c

ID AAL08306 standard; cDNA; 401 BP.

XX
AC AAL08306;

XX 07-DEC-2001 (first entry)

XX Human breast cancer expressed polynucleotide 763.

XX Human; breast cancer; cell marker; cytostatic; ss.

XX Homo sapiens.

XX

PN WO200151628-A2.
 XX 19-JUL-2001.
 PD 10-JAN-2001; 2001WO-US000798.
 XX 14-JAN-2000; 2000US-0176077P.
 PR 14-MAR-2000; 2000US-0189167P.
 PR 24-MAR-2000; 2000US-0192099P.
 PR 29-MAR-2000; 2000US-0193480P.
 PR 15-MAY-2000; 2000US-0205230P.
 PR 09-JUN-2000; 2000US-0211315P.
 PR 25-JUL-2000; 2000US-0220534P.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA Lillie J, Xu Y, Wang Y, Steinmann K;
 XX WPI; 2001-451856/48.
 DR New peptide useful as a marker for the diagnosis of breast cancer.
 PT Claim 1; Page 209; 3695pp; English.
 XX The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterizing treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity
 XX Sequence 401 BP; 92 A; 74 C; 81 G; 149 T; 0 U; 5 Other;
 SQ

Query Match 1.1%; Score 27; DB 4; Length 401;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2390 AATATAACCCCAAAAAAAAAAAAAAAAAA 2416
 DB 76 AATATAACCCCAAAAAAAAAAAAAAAAAA 50

RESULT 9
 AAL16849/c
 ID AAL16849 standard; cDNA; 471 BP.
 XX AAL16849;
 AC AAL16849;
 DT 07-DEC-2001 (first entry)
 DE Human breast cancer expressed polynucleotide 9306.
 XX Human; breast cancer; cell marker; cytostatic; ss.
 XX Homo sapiens.
 OS WO200151628-A2.
 PN 19-JUL-2001.
 PD 10-JAN-2001; 2001WO-US000798.
 XX 14-JAN-2000; 2000US-0176077P.
 PR 14-MAR-2000; 2000US-0189167P.
 PR 24-MAR-2000; 2000US-0192099P.
 PR 29-MAR-2000; 2000US-0193480P.
 PR 15-MAY-2000; 2000US-0205230P.
 PR 09-JUN-2000; 2000US-0211315P.
 PR 25-JUL-2000; 2000US-0220534P.
 XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX Lillie J, Xu Y, Wang Y, Steinmann K;
 XX WPI; 2001-451856/48.
 DR New peptide useful as a marker for the diagnosis of breast cancer.
 PT Claim 1; Page 1672; 3695pp; English.
 XX The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterizing treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity
 XX Sequence 471 BP; 115 A; 83 C; 101 G; 172 T; 0 U; 0 Other;
 SQ

Query Match 1.1%; Score 27; DB 4; Length 471;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2390 AATATAACCCCAAAAAAAAAAAAAAAAAA 2416
 DB 39 AATATAACCCCAAAAAAAAAAAAAAAAAA 13

RESULT 10
 ABL02795
 ID ABL02795 standard; cDNA; 2855 BP.
 XX ABL02795;
 AC ABL02795;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2867.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US009231.
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR P-PSDB; ABB58692.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.
 PT Claim 1; SEQ ID NO 2867; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention


```

CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (AB057737-
CC AB072072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2855 BP; 661 A; 750 C; 724 G; 720 T; 0 U; 0 Other;
Query Watch 1.1%; Score 26; DB 4; Length 2855;
Best Local Similarity 100.0%; Pred. NO. 3.6;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1937 GATATCAACAAGTTCCTGTGGATGGT 1962
DB 2213 GATATCAACAAGTTCCTGTGGATGGT 2238
|||||
RESULT 11
ADC37655
ID ADC37655 standard; cDNA; 3248 BP.
XX
AC ADC37655;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human angiomotin-like protein 1b encoding cDNA SEQ ID NO:4.
XX
KW human; angiomotin-like protein 1; AMLP1; cytostatic; gene therapy;
KW AMLP1b; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003037931-A2.
XX
PD 08-MAY-2003.
XX
PF 01-NOV-2002; 2002WO-US035129.
XX
PR 01-NOV-2001; 2001US-0334773P.
XX
PA (AMSH ) AMERSHAM BIOSCIENCES SV CORP.
XX
PI Shannon M, Phan T;
XX
XX WPI; 2003-430501/40.
DR P-PSDB; ADC37657.
XX
XX New isolated nucleic acid molecule encoding a human angiomotin-like
PT protein, useful for treating or preventing a disorder associated with
PT decreased or increased expression or activity of AMLP1.
XX
PS Claim 1; SEQ ID NO 4; 172pp; English.
XX
XX The present invention describes the human angiomotin-like protein 1
CC (AMLP1). human AMLP1 has cytostatic activity, and can be used in gene
CC therapy. The AMLP1 protein, nucleic acid molecules, antibodies, and
CC compositions of the present invention can be used for treating or
CC preventing a disorder associated with decreased or increased expression
CC or activity of AMLP1. The present sequence encodes human AMLP1b from the
CC present invention.
XX
SQ Sequence 3248 BP; 955 A; 865 C; 880 G; 548 T; 0 U; 0 Other;
Query Match 1.1%; Score 26; DB 10; Length 3248;
Best Local Similarity 100.0%; Pred. NO. 3.5;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2392 ATATACCCCAAAAAAAAAAAAAAAAAAAAA 2417
DB 3207 ATATACCCCAAAAAAAAAAAAAAAAAAAAA 3232
|||||
RESULT 12

```

```

PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 11603; 11750pp; English.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 11603; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 253 BP; 41 A; 69 C; 46 G; 97 T; 0 U; 0 Other;
Query Match 1.0%; Score 25; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. NO. 13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2393 TAAACCCAAAAA 2417
Db 29 TAAACCCAAAAA 5
RESULT 14
ABV61368/c
ID ABV61368 standard; cDNA; 282 BP.
XX
AC ABV61368;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 61359.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 11641; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 282 BP; 69 A; 66 C; 61 G; 86 T; 0 U; 0 Other;
Query Match 1.0%; Score 25; DB 5; Length 282;
Best Local Similarity 100.0%; Pred. NO. 13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2393 TAAACCCAAAAA 2417
Db 42 TAAACCCAAAAA 18
RESULT 15
ABV43386/c
ID ABV43386 standard; cDNA; 340 BP.
XX
AC ABV43386;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 43377.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.

```

XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX
PS Claim 1; Page 8649; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 340 BP; 97 A; 46 C; 63 G; 134 T; 0 U; 0 Other;

Query Match 1.0%; Score 25; DB 5; Length 340;
Best Local Similarity 100.0%; Pred.No. 13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2393 TAAACCCAAAAA 2417
Db 84 TAAACCCAAAAA 60

Search completed: December 15, 2004, 15:47:00
Job time : 1627 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 10:34:04 ; Search time 253.148 Seconds
(without alignments)
6786.440 Million cell updates/sec

Title: US-10-028-384-7

Perfect score: 2417

Sequence: 1 tctaaagcgaagaatgtgtcg.....ccaaaaaAAAAAAAAAAAA 2417

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	889	36.8	900	4	US-09-270-767-12331
2	755	31.2	1660	4	US-09-270-767-158
3	755	31.2	1660	4	US-09-270-767-15440
C 4	503	20.8	560	4	US-09-270-767-1318
C 5	503	20.8	560	4	US-09-270-767-16600
C 6	299	12.4	1660	4	US-09-270-767-158
C 7	299	12.4	1660	4	US-09-270-767-15440
8	133	5.5	133	4	US-09-270-767-28082
9	26	1.1	914	4	US-09-270-767-12856
10	24	1.0	632	3	US-09-385-982-499
11	24	1.0	660	1	US-08-555-678-41
12	24	1.0	1060	4	US-08-997-685A-7
13	24	1.0	1182	3	US-09-461-637-261
14	24	1.0	1391	1	US-08-261-652-1
15	24	1.0	1391	5	PCR-US95-07752-1
16	24	1.0	1720	4	US-09-823-038A-41
17	24	1.0	4137	3	US-09-499-964-2
18	24	1.0	4302	3	US-09-245-281-38
19	24	1.0	4302	4	US-09-207-359B-38
20	24	1.0	4302	4	US-09-340-620A-38
21	24	1.0	4302	4	US-09-865-364-38
22	23	1.0	132	4	US-09-621-976-16776
23	23	1.0	134	4	US-09-621-976-16780
24	23	1.0	135	4	US-09-621-976-16778
25	23	1.0	139	4	US-09-621-976-16785
26	23	1.0	140	4	US-09-621-976-16783
27	23	1.0	140	4	US-09-621-976-16786

28	23	1.0	147	4	US-09-270-767-13687	Sequence 13687, Ap
29	23	1.0	147	4	US-09-270-767-18969	Sequence 18969, A
30	23	1.0	157	4	US-09-621-976-16781	Sequence 16781, A
31	23	1.0	190	4	US-09-621-976-16784	Sequence 16784, A
32	23	1.0	383	4	US-09-621-976-12067	Sequence 12067, A
33	23	1.0	383	4	US-09-621-976-12237	Sequence 12237, A
34	23	1.0	383	4	US-09-621-976-13563	Sequence 13563, A
35	23	1.0	453	4	US-09-621-976-2916	Sequence 2916, Ap
36	23	1.0	1071	4	US-09-205-258-118	Sequence 118, App
37	23	1.0	1133	4	US-09-205-258-236	Sequence 236, App
38	23	1.0	1316	4	US-09-370-838-30	Sequence 30, Appl
39	23	1.0	1316	4	US-09-854-133-30	Sequence 30, Appl
40	23	1.0	2030	3	US-08-706-216-3	Sequence 30, Appl
41	23	1.0	2030	4	US-09-650-284B-3	Sequence 3, Appl1
42	23	1.0	2254	4	US-09-138-277C-2	Sequence 2, Appl1
43	23	1.0	3467	4	US-09-743-492A-1	Sequence 1, Appl1
C 44	23	1.0	5455	4	US-10-204-708-33	Sequence 33, Appl
45	23	1.0	7702	4	US-09-023-655-1336	Sequence 1336, Ap

ALIGNMENTS

RESULT 1

US-09-270-767-12331
; Sequence 12331, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 12331
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-12331

Query Match 36.8%; Score 889; DB 4; Length 900;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 889; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TCTAAGCGAAGATGTGCTGCTTCAGATTCAGATTCGATTAATTTTCGATTTACTGGCTG	60
DB	12	TCTAAGCGAAGATGTGCTGCTTCAGATTCAGATTCGATTAATTTTCGATTTACTGGCTG	71
QY	61	GAATTTGGGACATGAATCGGACGCGAAGATGCTGAACAGCAAGTGTGCTGGCTACAGCAG	120
DB	72	GAATTTGGGACATGAATCGGACGCGAAGATGCTGAACAGCAAGTGTGCTGGCTACAGCAG	131
QY	121	CCTAATACCTTCGCCATCTGCTTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTT	180
DB	132	CCTAATACCTTCGCCATCTGCTTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTT	191
QY	181	CGCGTCATCGTTTCGAGTCGATTATCCATGATTTGATCCGCTGGCTGGCTACACCGGC	240
DB	192	CGCGTCATCGTTTCGAGTCGATTATCCATGATTTGATCCGCTGGCTGGCTACACCGGC	251
QY	241	CACCGCTACATGTCGAGAAATGGTTGGTACAATCTTCTCAACTGGTTCGACGAGCGGC	300
DB	252	CACCGCTACATGTCGAGAAATGGTTGGTACAATCTTCTCAACTGGTTCGACGAGCGGC	311
QY	301	ATGCTATCCGCTCGCGAGGATTTGGCGGTACCGTCTATCCCGGCTGATGATTACGTC	360
DB	312	ATGCTATCCGCTCGCGAGGATTTGGCGGTACCGTCTATCCCGGCTGATGATTACGTC	371
QY	361	CGCGGAATCAATGGCTGTGCAACGTAATCAACATACCGGTCCATTCGTCGATTCG	420
DB	372	CGCGGAATCAATGGCTGTGCAACGTAATCAACATACCGGTCCATTCGTCGATTCG	431

Query Match	20.8%	Score 503;	DB 4;	Length 560;
Best Local Similarity	99.8%	Pred. No. 7.2e-208;		
Matches 553;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1107	TCATTGCATCCGTTGTCGGAGCATCAGCCACACATGTTCTCGTTCCTCTTTGATCTGC	1166	
Db	560	TCATTGCATCCGTTGTCGGAGCATCAGCCACCACTTGGTTCTCGTTCCTTTGATCTGC	501	
QY	1167	ACATCCCTGGTGTGCGCTTCCCATGGGAGTGTGCTACTGCATCAAGCAGATCAACGACG	1226	
Db	500	ACATCATGTTGTGCGCTTCCCATGGGAGTGTGCTACTGCATCAAGCAGATCAACGACG	441	
QY	1227	ACGCGGTTTTTCGTGGTGTCTGTACGCCATCAGTCGGGTTTACTTCGCTGGTGTGATGTGC	1286	
Db	440	ACGCGGTTTTTCGTGGTGTCTGTACGCCATCAGTTCGCGTTTACTTCGCTGGTGTGATGTGC	381	
QY	1287	GTTTTGATGTTTACACCCCTCACGCCGGTGGTGTGATGCTGCGCGAGTGGCCCTTTTCGGGAC	1346	

[illegible]

QY 1467 GCAAGCTGAAGCATCTACTAAGCATGATGCCAGCAGAGTACTGGCTCAGCTCCAAACC 1526
DB 200 GCAAGCTGAAGCATCTACTAAGCATGATGCCAGCAGAGTACTGGCTCAGCTCCAAACC 141
QY 1527 TGAAGAGTATTGTTATTGTCGGTCTTAATGCTGTGATGATGTTTCGCTCCACTGCA 1586
DB 140 TGAAGAGTATTGTTATTGTCGGTCTTAATGCTGTGATGATGTTTCGCTCCACTGCA 81
QY 1587 CGTGGGTGACCAAGTCCCTACTCCAGTCCCTCCATGCTCTGGGCTTTCCACACAGTC 1646
DB 80 CGTGGGTGACCAAGTCCCTACTCCAGTCCCTCCATGCTCTGGGCTTTCCACACAGTC 21
QY 1647 AAGATGATCCCGC 1660
DB 20 AAGATGATCCCGC 7

RESULT 6

US-09-270-767-158/c
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

Query Match 12.4%; Score 299; DB 4; Length 1660;
Best Local Similarity 99.1%; Pred. No. 1e-119;
Matches 549; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1760 GCGGGAATGGCAACAGACGACCTAGTGGATAATATACGTGGAAACAATAGTCACATA 1819
DB 793 GCGGGAATGGCAACAGACGACCTAGTGGATAATATACGTGGAAACAATAGTCACATA 734
QY 1820 GCGCTGTTGGCAAGCAATGTTCTTAACCGAGAGAGTCTTACGAATATATGACATCT 1879
DB 733 GCGCTGTTGGCAAGCAATGTTCTTAACCGAGAGAGTCTTACGAATATATGACATCT 674
QY 1880 CTTCAGCTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGAT 1939
DB 673 CTTCAGCTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGAT 614
QY 1940 ATCAACAAGTTCCTGTGGATGTCGGAATTTAGCTACTACAGATTCGGGAATTTGAAGTTGAC 1999
DB 613 ATCAACAAGTTCCTGTGGATGTCGGAATTTAGCTACTACAGATTCGGGAATTTGAAGTTGAC 554
QY 2000 GAAAGCCGATTAATCTTACCGACCGCGGTGAATTCAGGCTAGATCCGAGGAGTCTCCGGCC 2059
DB 553 GAAAGCCGATTAATCTTACCGACCGCGGTGAATTCAGGCTAGATCCGAGGAGTCTCCGGCC 494
QY 2060 CTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGAATTTGAAGTTGAC 2119
DB 493 CTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGAATTTGAAGTTGAC 434
QY 2120 TACAGAGTCCATCTGGATATGATCGACACGTAAGCGCGTCAATTCGGGAATTAAGGACTTC 2179
DB 433 TACAGAGTCCATCTGGATATGATCGACACGTAAGCGCGTCAATTCGGGAATTAAGGACTTC 374
QY 2180 GATCTGACCTTACTCGAGAGGCGCTTACACAGACACTGGCTTGTTCGCTATTAAG 2239
DB 373 GATCTGACCTTACTCGAGAGGCGCTTACACAGACACTGGCTTGTTCGCTATTAAG 314
QY 2240 GTGAAGAGCCGCTAGTTCATTAAGACCATCTACTGAAGACCAAGGAGAGAAAGATTCCT 2299

RESULT 8

US-09-270-767-28082
; Sequence 28082, Application US/09270767

DB 313 GTGAAGAGCCGCTAGTTCATATAGACCATCATCTGAAGCAAGGAGAGACGATTCCT 254
QY 2300 CCAGCAAACTTCAT 2313
DB 253 CCAGCAAACTTCAT 240
RESULT 7
US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Query Match 12.4%; Score 299; DB 4; Length 1660;
Best Local Similarity 99.1%; Pred. No. 1e-119;
Matches 549; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1760 GCGGGAATGGCAACAGACGACCTAGTGGATAATATACGTGGAAACAATAGTCACATA 1819
DB 793 GCGGGAATGGCAACAGACGACCTAGTGGATAATATACGTGGAAACAATAGTCACATA 734
QY 1820 GCGCTGTTGGCAAGCAATGTTCTTAACCGAGAGAGTCTTACGAATATATGACATCT 1879
DB 733 GCGCTGTTGGCAAGCAATGTTCTTAACCGAGAGAGTCTTACGAATATATGACATCT 674
QY 1880 CTTCAGCTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGAT 1939
DB 673 CTTCAGCTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGAT 614
QY 1940 ATCAACAAGTTCCTGTGGATGTCGGAATTTAGCTACTACAGATTCGGGAATTTGAAGTTGAC 1999
DB 613 ATCAACAAGTTCCTGTGGATGTCGGAATTTAGCTACTACAGATTCGGGAATTTGAAGTTGAC 554
QY 2000 GAAAGCCGATTAATCTTACCGACCGCGGTGAATTCAGGCTAGATCCGAGGAGTCTCCGGCC 2059
DB 553 GAAAGCCGATTAATCTTACCGACCGCGGTGAATTCAGGCTAGATCCGAGGAGTCTCCGGCC 494
QY 2060 CTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGAATTTGAAGTTGAC 2119
DB 493 CTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGAATTTGAAGTTGAC 434
QY 2120 TACAGAGTCCATCTGGATATGATCGACACGTAAGCGCGTCAATTCGGGAATTAAGGACTTC 2179
DB 433 TACAGAGTCCATCTGGATATGATCGACACGTAAGCGCGTCAATTCGGGAATTAAGGACTTC 374
QY 2180 GATCTGACCTTACTCGAGAGGCGCTTACACAGACACTGGCTTGTTCGCTATTAAG 2239
DB 373 GATCTGACCTTACTCGAGAGGCGCTTACACAGACACTGGCTTGTTCGCTATTAAG 314
QY 2240 GTGAAGAGCCGCTAGTTCATTAAGACCATCTACTGAAGACCAAGGAGAGAAAGATTCCT 2299
DB 313 GTGAAGAGCCGCTAGTTCATTAAGACCATCTACTGAAGACCAAGGAGAGAAAGATTCCT 254
QY 2300 CCAGCAAACTTCAT 2313
DB 253 CCAGCAAACTTCAT 240

Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28082
; LENGTH: 133
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-28082

Query Match 5.5%; Score 133; DB 4; Length 133;
Best Local Similarity 100.0%; Pred. No. 6.5e-48;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 757 TATGGCAGGTACTCGCGCGTCTGCTGACGAGCTACAGCCTTCTACATCCTGGGACT 816
Db 1 TATGGCAGGTACTCGCGCGTCTGCTGACGAGCTACAGCCTTCTACATCCTGGGACT 60

QY 817 GCTGTTCTCCATGCAGATCCCTTCTGTTGGATTCCACCGATACGCCACGAGTGAACACAT 876
Db 61 GCTGTTCTCCATGCAGATCCCTTCTGTTGGATTCCACCGATACGCCACGAGTGAACACAT 120

QY 877 GCTGCGCTGGGA 889
Db 121 GCTGCGCTGGGA 133

RESULT 9
US-09-270-767-12856
; Sequence 12856, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12856
; LENGTH: 914
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-12856

Query Match 1.1%; Score 26; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1937 GATATCAACAAGTCTCTGTGGATGTT 1962
Db 272 GATATCAACAAGTCTCTGTGGATGTT 297

RESULT 10
US-09-385-982-499
; Sequence 499, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393

; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 499
; LENGTH: 632
; TYPE: DNA
; ORGANISM: *Homo sapiens*
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(632)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-499

Query Match 1.0%; Score 24; DB 3; Length 632;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2394 AAACCCAAAAA 2417
Db 505 AAACCCAAAAA 528

RESULT 11
US-08-555-678-41
; Sequence 41, Application US/08555678
; Patent No. 5763174
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,678
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/280,443
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,459
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49DUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 660 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-555-678-41

Query Match 1.0%; Score 24; DB 1; Length 660;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2394 AAACCCAAAAA 2417
Db 114 AAACCCAAAAA 137

RESULT 12

US-08-997-685A-7/c
; Sequence 7, Application US/08997685A
; Patent No. 6551821
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University
; APPLICANT: Kandel, Eric
; TITLE OF INVENTION: Brain Cyclic Nucleotide Gated Ion Channel and Uses Thereof
; FILE REFERENCE: 0575/54806
; CURRENT APPLICATION NUMBER: US/08/997,685A
; CURRENT FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: mouse;
US-08-997-685A-7

Query Match 1.0%; Score 24; DB 4; Length 1060;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2394 AAACCCAAAAA 2417
Db 28 AAACCCAAAAA 5

RESULT 13

US-09-461-697-261
; Sequence 261, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Furham, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 261
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-261

Query Match 1.0%; Score 24; DB 3; Length 1182;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2394 AAACCCAAAAA 2417
Db 1150 AAACCCAAAAA 1173

RESULT 14

US-08-261-662-1
; Sequence 1, Application US/08261662
; Patent No. 5523227
; GENERAL INFORMATION:
; APPLICANT: Bram, Richard J
; APPLICANT: Crabtree, Gerald R
; TITLE OF INVENTION: CELL CALCIUM REGULATION AND ITS USE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,662
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 43s
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-59392/BIR STAN-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1391 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 37..927
US-08-261-662-1

Query Match 1.0%; Score 24; DB 1; Length 1391;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2394 AAACCCAAAAA 2417
Db 1343 AAACCCAAAAA 1366

RESULT 15

PCI-US95-07752-1
; Sequence 1, Application PC/TUS9507752
; GENERAL INFORMATION:
; APPLICANT: The Board of Trustees of the Leland Stanford Junior
; APPLICANT: University
; TITLE OF INVENTION: CELL CALCIUM REGULATION AND ITS USE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07752
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A-59392/DIR STAN-167
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1391 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 37..927
PCT-US95-07752-1

Query Match 1.0%; Score 24; DB 5; Length 1391;
Best Local Similarity 100.0%; Pred.No. 0.75;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2394 AAACCCAAAAAAAAAAAAAAAAAAAA 2417
Db 1343 AAACCCAAAAAAAAAAAAAAAAAAAA 1366

Search completed: December 16, 2004, 07:21:31
Job time : 255.148 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2004, 00:55:34 ; Search time 2138.68 Seconds
(without alignments)
6239.285 Million cell updates/sec

Title: US-10-028-384-7

Perfect score: 2417

Sequence: 1 tctagcgaagaatgtctg.....ccaaaaaataaaaaaa 2417

Scoring table: OLIGO_NUC

Gapop_60.0 , Gapext 60.0

Searched: 4093002 seqs, 2760418825 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8186004

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2417	100.0	2417	15	US-10-028-384-7
C 2	28	1.2	442	11	US-09-969-034-3604
C 3	26	1.1	354	18	US-10-425-115-67624
C 4	26	1.1	395	14	US-10-065-543-1610
5	26	1.1	348	18	US-10-494-343-4
6	25	1.0	153	16	US-10-242-535A-53154
7	25	1.0	153	16	US-10-085-783A-53154
8	25	1.0	293	16	US-10-242-535A-14946
9	25	1.0	293	16	US-10-085-783A-14946
C 10	25	1.0	354	16	US-10-131-827-8437
11	25	1.0	416	16	US-10-242-535A-44394
12	25	1.0	416	16	US-10-085-783A-44394

C	13	25	1.0	488	18	US-10-425-115-78691	Sequence 78691, A
	14	25	1.0	513	10	US-09-814-353-2031	Sequence 2031, Ap
	15	25	1.0	513	10	US-09-814-353-8377	Sequence 8377, Ap
C	16	25	1.0	572	16	US-10-191-803-934	Sequence 934, App
	17	25	1.0	583	10	US-09-814-353-14761	Sequence 14761, A
	18	25	1.0	592	18	US-10-425-115-114851	Sequence 114851, A
	19	25	1.0	600	16	US-10-424-599-99562	Sequence 99562, A
	20	25	1.0	659	18	US-10-363-345A-14705	Sequence 14705, A
	21	25	1.0	659	18	US-10-363-345A-14706	Sequence 14706, A
C	22	25	1.0	679	18	US-10-425-115-47797	Sequence 47797, A
	23	25	1.0	1023	9	US-09-804-682-24	Sequence 24, Appl
	24	25	1.0	1033	9	US-09-804-682-23	Sequence 23, Appl
	25	25	1.0	2335	15	US-10-341-434-237	Sequence 237, App
C	26	25	1.0	5244	18	US-10-473-126-312	Sequence 312, App
	27	25	1.0	8044	16	US-10-221-714A-347	Sequence 347, App
C	28	25	1.0	23611	17	US-10-322-281-61	Sequence 61, Appl
	29	24	1.0	139	17	US-10-021-323-5739	Sequence 5739, Ap
C	30	24	1.0	143	10	US-09-814-353-18749	Sequence 18749, A
	31	24	1.0	143	16	US-10-242-535A-51365	Sequence 51365, A
	32	24	1.0	143	16	US-10-085-783A-51365	Sequence 51365, A
C	33	24	1.0	167	10	US-09-814-353-2461	Sequence 2461, Ap
C	34	24	1.0	167	10	US-09-814-353-8798	Sequence 8798, Ap
C	35	24	1.0	175	9	US-09-983-965-2999	Sequence 2999, Ap
C	36	24	1.0	180	16	US-10-424-599-69408	Sequence 69408, A
C	37	24	1.0	200	11	US-09-969-034-321	Sequence 321, Appl
	38	24	1.0	224	9	US-09-784-846-58	Sequence 58, Appl
	39	24	1.0	224	14	US-10-091-483-58	Sequence 58, Appl
C	40	24	1.0	229	16	US-10-062-674-971	Sequence 971, App
	41	24	1.0	243	16	US-10-242-535A-16977	Sequence 16977, A
	42	24	1.0	243	16	US-10-085-783A-16977	Sequence 16977, A
C	43	24	1.0	244	18	US-10-425-115-10205	Sequence 10205, A
C	44	24	1.0	247	18	US-10-425-115-148698	Sequence 148698, A
C	45	24	1.0	251	18	US-10-425-115-166919	Sequence 166919, A

ALIGNMENTS

RESULT 1

US-10-028-384-7
; Sequence 7, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 2417
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AF132552
; DATABASE ENTRY DATE: 1999-04-27
; RELEVANT RESIDUES: (1)..(2417)
US-10-028-384-7

Query Match 100.0%; Score 2417; DB 15; Length 2417;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAGCGAAGAATGTCTGCTTGCATTCAGATCGGTATAATTTTCGAGTTACTGGCTG 60
DB 1 TCTAGCGAAGAATGTCTGCTTGCATTCAGATCGGTATAATTTTCGAGTTACTGGCTG 60
QY 61 GAATTTGGACATGAATCGGACCGCAAGATCCTGAACAGCAAGTGGCTGGCTACAGCAG 120
DB 61 GAATTTGGACATGAATCGGACCGCAAGATCCTGAACAGCAAGTGGCTGGCTACAGCAG 120

121 CCTAATCACTTCGCGCATCTGCTAATCGCTGGCTGGCGGAGTTTCTCTCGCTCTT 180
121 CCTAATCACTTCGCGCATCTGCTAATCGCTGGCTGGCGGAGTTTCTCTCGCTCTT 180
181 CGCGTCAATCGTTTCAGTCGATTAATCCATGAGTTTGTATCCGTGGTTCAATACCGGGC 240
181 CGCGTCAATCGTTTCAGTCGATTAATCCATGAGTTTGTATCCGTGGTTCAATACCGGGC 240
241 CACCGCTTACATGCTGCGAGATGTTGGTACAACTTCTCAACTGTTTCAAGCGGGC 300
241 CACCGCTTACATGCTGCGAGATGTTGGTACAACTTCTCAACTGTTTCAAGCGGGC 300
301 ATGGTATCCGCTCGGCGAGATGTTGGCGGTATCGCTATCCCGGGCTCATGATTAACGTC 360
301 ATGGTATCCGCTCGGCGAGATGTTGGCGGTATCGCTATCCCGGGCTCATGATTAACGTC 360
361 CGCGGGAATCCATGCTGCTGCACTGCTAATCAATCAATCAATCAATCAATCAATCAATCA 420
361 CGCGGGAATCCATGCTGCTGCACTGCTAATCAATCAATCAATCAATCAATCAATCAATCA 420
421 CGTGTTCCTGGCGCGATCTTCACTGGCTGACCTCCATCCACCTACCTGCTGACCAA 480
421 CGTGTTCCTGGCGCGATCTTCACTGGCTGACCTCCATCCACCTACCTGCTGACCAA 480
481 GGAGTGTGCTCGCGGGCGCGCTCTTTCGCGCGAGCTTTCGCGCGAGCTTTCGCGCGAGCT 540
481 GGAGTGTGCTCGCGGGCGCGCTCTTTCGCGCGAGCTTTCGCGCGAGCTTTCGCGCGAGCT 540
541 CTACATCAGTAGTGGCTGCTGCTACATGCTTCTACATGCTTCTGCTGCTGCTGCTGCTGCT 600
541 CTACATCAGTAGTGGCTGCTGCTACATGCTTCTACATGCTTCTGCTGCTGCTGCTGCTGCT 600
601 GCAGTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
601 GCAGTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
661 CGACGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
661 CGACGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
721 CAACCTGATACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
721 CAACCTGATACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
781 GCTGACGAGCTACAGCACTTCTACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
781 GCTGACGAGCTACAGCACTTCTACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
841 CGTGGGATTCACCGGATAGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
841 CGTGGGATTCACCGGATAGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
901 CCTATGCGCTGGCGCACTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
901 CCTATGCGCTGGCGCACTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
961 GAAGCTGTTTCACTGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
961 GAAGCTGTTTCACTGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
1021 GCTCACCATGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
1021 GCTCACCATGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
1081 TGGCTACCGCAAGATTCACATTCCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
1081 TGGCTACCGCAAGATTCACATTCCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
1141 TTGGTTCCTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
1141 TTGGTTCCTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

1201 GTACTGCTCAAGCAGATCAACGACGCGGCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
1201 GTACTGCTCAAGCAGATCAACGACGCGGCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
1261 GGTTTACTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
1261 GGTTTACTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
1321 GCTGGCGGAGTGGCTTTTTCGGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
1321 GCTGGCGGAGTGGCTTTTTCGGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
1381 GCGAATGGGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
1381 GCGAATGGGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
1441 GAAGAAGCCTGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1500
1441 GAAGAAGCCTGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1500
1501 GCAGGATCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1501 GCAGGATCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1561 GTTGATGATGTTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
1561 GTTGATGATGTTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
1621 CATTTGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
1621 CATTTGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
1681 AGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
1681 AGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
1741 GGATTCAGGATACAGATAGCGGAAATGGCAACAGAACAGAACAGAACAGAACAGAACAGAAC 1800
1741 GGATTCAGGATACAGATAGCGGAAATGGCAACAGAACAGAACAGAACAGAACAGAACAGAAC 1800
1801 GTGGAACATAGTACATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
1801 GTGGAACATAGTACATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
1861 CTACGAAATTAATGACATCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
1861 CTACGAAATTAATGACATCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
1921 CGGCTATTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
1921 CGGCTATTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
1981 GCATCCAAAGGACATTAAGGAAAGGATTTACTTTACCGACCGGCTGAAATTCAGGGTAGA 2040
1981 GCATCCAAAGGACATTAAGGAAAGGATTTACTTTACCGACCGGCTGAAATTCAGGGTAGA 2040
2041 TGCAGAGGCTGCTCGGCGCTGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
2041 TGCAGAGGCTGCTCGGCGCTGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
2101 CGGGGAATTAAGGATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 2160
2101 CGGGGAATTAAGGATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 2160
2161 CATTTGGATTAAGGATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 2220
2161 CATTTGGATTAAGGATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 2220
2221 GCTTGTTCGCTCTATAGGTTGAAGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
2221 GCTTGTTCGCTCTATAGGTTGAAGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
2281 CRAAGGAGAGACGATTTCTCCAGCAAACTTCTCATTTCCAGGAAAGAACTCTTAAGCGCTGCA 2340

Db 2281 CAAGGAGAGACGATTCCTCCAGCAAACTTCATTTCCGAGAGAACTCTAAGCTCGCAA 2340
QY 2341 GGGCTACATACGACACCGGTTGTTGTTAAAGGGAACGAACTTGAAATAAACCCA 2400
Db 2341 GGGCTACATACGAAACCGGTTGTTGTTAAAGGGAACGAACTTGAAATAAACCCA 2400
QY 2401 AAAAAAAAAAAAAAAAAA 2417
Db 2401 AAAAAAAAAAAAAAAAAA 2417

RESULT 2

US-09-969-034-3604/c
; Sequence 3604, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwivedi, Poorima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; TITLE OF INVENTION: Expressed in Cancer Tissue
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3604
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 89, 220, 278, 287, 337, 350, 378, 391, 404, 406, 422, 432,
; LOCATION: 434, 435, 442
; OTHER INFORMATION: n = A,T,C or G
US-09-969-034-3604

Query Match 1.2% Score 28; DB 11; Length 442;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2390 AAATAAACCCCAAAAAAAAAAAAAAAAAA 2417
Db 34 AAATAAACCCCAAAAAAAAAAAAAAAAAA 7

RESULT 3

US-10-425-115-67624/c
; Sequence 67624, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 67624
; LENGTH: 354
; TYPE: DNA
; ORGANISM: Zea mays

; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_16166C.1
US-10-425-115-67624

Query Match 1.1% Score 26; DB 18; Length 354;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2392 ATAAACCCCAAAAAAAAAAAAAAAAAA 2417
Db 79 ATAAACCCCAAAAAAAAAAAAAAAAAA 54

RESULT 4

US-10-066-543-1610/c
; Sequence 1610, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Joseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1610
; LENGTH: 395
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 87, 252, 284, 315, 391
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-1610

Query Match 1.1% Score 26; DB 14; Length 395;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2392 ATAAACCCCAAAAAAAAAAAAAAAAAA 2417
Db 40 ATAAACCCCAAAAAAAAAAAAAAAAAA 15

RESULT 5

US-10-494-343-4
; Sequence 4, Application US/10494343
; Publication No. US20040248138A1
; GENERAL INFORMATION:
; APPLICANT: Shannon, Mark
; APPLICANT: Phan, Thuyvy
; TITLE OF INVENTION: HUMAN AGIOMOTIN-LIKE PROTEIN 1
; FILE REFERENCE: PB0184
; CURRENT APPLICATION NUMBER: US/10/494,343
; CURRENT FILING DATE: 2004-04-30
; PRIOR APPLICATION NUMBER: US to be assigned
; PRIOR FILING DATE: to be assigned
; PRIOR APPLICATION NUMBER: PCT/US2002/035129
; PRIOR FILING DATE: 2002-11-01
; PRIOR APPLICATION NUMBER: US 60/334,773
; PRIOR FILING DATE: 2001-11-01
; NUMBER OF SEQ ID NOS: 870

us-10-028-384-7.Oligo.rnpb

Thu Dec 16 16:25:11 2004

```
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 4
; LENGTH: 3248
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-494-343-4

Query Match      1.1%; Score 26; DB 18; Length 3248;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2392 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 3207 ATAACCCCAAAAAAAAAAAAAAAAAAAAAA 3232

RESULT 6
US-10-242-535A-53154
; Sequence 53154, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: Liow, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 53154
; LENGTH: 153
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (33)..(33)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-53154

Query Match      1.0%; Score 25; DB 16; Length 153;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 122 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 146

RESULT 7
US-10-085-783A-14946
; Sequence 53154, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liow, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28

Query Match      1.0%; Score 25; DB 16; Length 153;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 122 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 146

RESULT 8
US-10-242-535A-14946
; Sequence 14946, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liow, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14946
; LENGTH: 293
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-14946

Query Match      1.0%; Score 25; DB 16; Length 293;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 260 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 284

RESULT 9
US-10-085-783A-14946
; Sequence 14946, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liow, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
```


; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14946
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-14946

Query Match 1.0%; Score 25; DB 16; Length 293;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2393 TAAACCCCAAAAAAAAAAAAAA 2417
Db 260 TAAACCCCAAAAAAAAAAAAAA 284

RESULT 10

US-10-131-827-8437/c
; Sequence 8437, Application US/10131827
; Publication No. US20040009479A1
; GENERAL INFORMATION:
; APPLICANT: Wohlgemuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 50661200120
; CURRENT APPLICATION NUMBER: US/10/131.827
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8437
; LENGTH: 354
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-827-8437

Query Match 1.0%; Score 25; DB 16; Length 354;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2393 TAAACCCCAAAAAAAAAAAAAA 2417
Db 40 TAAACCCCAAAAAAAAAAAAAA 16

RESULT 11

US-10-242-535A-44394
; Sequence 44394, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 44394
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-44394

Query Match 1.0%; Score 25; DB 16; Length 416;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2393 TAAACCCCAAAAAAAAAAAAAA 2417
Db 392 TAAACCCCAAAAAAAAAAAAAA 416

RESULT 12

US-10-085-783A-44394
; Sequence 44394, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 44394
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-44394

Query Match 1.0%; Score 25; DB 16; Length 416;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2393 TAAACCCCAAAAAAAAAAAAAA 2417
Db 392 TAAACCCCAAAAAAAAAAAAAA 416

RESULT 13

US-10-425-115-78691/c
; Sequence 78691, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 78691
; LENGTH: 488
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(488)
; OTHER INFORMATION: unsure at all n locations

us-10-028-384-7.Oligo.rnpb

Thu Dec 16 16:25:11 2004

```

;
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_171793C.1
; US-10-425-115-78691
;
; Query Match
; Best Local Similarity 100.0%; Score 25; DB 18; Length 488;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 2392 TAAACCCCAAAAAAAAAAAAAAAAAA 2416
; |
; Db 179 ATAACCCCAAAAAAAAAAAAAAAAAA 155
;
; RESULT 14
; US-09-814-353-2031
; Sequence 2031, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lee, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2031
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 110
; OTHER INFORMATION: n = A,T,C or G
; US-09-814-353-8377
;
; Query Match
; Best Local Similarity 100.0%; Score 25; DB 10; Length 513;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAA 2417
; |
; Db 486 TAAACCCCAAAAAAAAAAAAAAAAAA 510
;
; Search completed: December 16, 2004, 16:17:35
; Job time : 2139.88 secs

```

```

;
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_171793C.1
; US-10-425-115-78691
;
; Query Match
; Best Local Similarity 100.0%; Score 25; DB 18; Length 488;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 2392 TAAACCCCAAAAAAAAAAAAAAAAAA 2416
; |
; Db 179 ATAACCCCAAAAAAAAAAAAAAAAAA 155
;
; RESULT 14
; US-09-814-353-2031
; Sequence 2031, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lee, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2031
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 110
; OTHER INFORMATION: n = A,T,C or G
; US-09-814-353-2031
;
; Query Match
; Best Local Similarity 100.0%; Score 25; DB 10; Length 513;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAA 2417
; |
; Db 486 TAAACCCCAAAAAAAAAAAAAAAAAA 510
;
; RESULT 15
; US-09-814-353-8377
; Sequence 8377, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lee, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 10:16:54 ; Search time 11093.6 Seconds
(without alignments)
7939.229 Million cell updates/sec

Title: US-10-028-384-7
Perfect score: 2417
Sequence: 1 tctagcagaagtgtctg.....cccaaaaaaaaaaaaaaa 2417

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 32822875 seqs, 18219865908 residues

Word size : 0

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	734	30.4	810	2	BF502026	BF502026 AT17637.5
2	691	28.2	681	7	CK659033	CK659033 LP16462.5
3	664	27.5	701	2	BF503420	BF503420 AT19406.5
4	658	27.2	666	4	BI357074	BI357074 RE43425.5
5	658	27.2	668	4	BI227902	BI227902 RE25288.5
6	657	27.2	715	7	CK657451	CK657451 LP23759.5
7	652	27.0	652	4	BI171940	BI171940 RE13841.5
8	644	26.6	644	4	BG636414	BG636414 SD14123.5
9	644	26.6	654	4	BI374334	BI374334 RE61893.5
10	641	26.5	641	1	AI257750	AI257750 LP06212.5
11	641	26.5	679	7	CK657649	CK657649 LP24119.5
12	639	26.4	690	4	BG640860	BG640860 SD12024.5
13	635	26.3	642	4	BI173041	BI173041 RE15774.5
14	623	25.8	635	1	AI135629	AI135629 GH13452.5
15	615	25.4	655	2	BF496296	BF496296 AT10060.5
16	615	25.4	663	4	BG641172	BG641172 SD12448.5
17	612	25.3	615	4	BI364613	BI364613 RE49773.5
18	607	25.1	655	4	BI484774	BI484774 RE67926.5
19	600	24.8	600	4	BI163592	BI163592 RE03026.5
20	588	24.3	630	1	AI295381	AI295381 LP08987.5
21	586	24.2	616	2	BF504378	BF504378 AT05729.5
22	585	24.2	585	1	AI133998	AI133998 GH1337.5
23	576	23.8	585	4	BI242003	BI242003 RE39217.5
24	567	23.5	649	4	BI162865	BI162865 RE02124.5

25	566	23.4	598	4	BI170396	BI170396 RE11825.5
26	564	23.3	658	4	BI374189	BI374189 RE61718.5
27	551	22.8	564	1	AI542327	AI542327 SD08615.5
28	549	22.7	735	1	AA949890	AA949890 LD29946.5
29	548	22.7	555	4	BI370013	BI370013 RE56375.5
30	538	22.3	682	7	CK659064	CK659064 LP16523.5
31	535	22.1	646	4	BI170481	BI170481 RE11944.5
32	529	21.9	554	4	BI167839	BI167839 RE08321.5
33	524	21.7	564	4	BI635692	BI635692 SD17119.5
34	511	21.1	525	2	BF502679	BF502679 AT18477.5
35	507	21.0	527	4	BI586931	BI586931 RH27159.5
36	507	21.0	539	1	AI107144	AI107144 GH06456.5
37	502	20.8	539	4	BI373819	BI373819 RE61243.5
38	495	20.5	495	1	AI533215	AI533215 SD04365.5
39	492	20.4	502	4	BI237763	BI237763 RE34052.5
40	480	19.9	480	1	AI514952	AI514952 LD46592.5
41	464	19.2	464	1	AI297771	AI297771 LP12143.5
42	458	18.9	478	1	AI533035	AI533035 SD04713.5
43	454	18.8	499	4	BI173583	BI173583 RE16531.5
44	441	18.2	445	4	BI579570	BI579570 RE73627.5
45	441	18.2	467	4	BI364972	BI364972 RE50181.5

ALIGNMENTS

RESULT 1
BF502026
LOCUS
DEFINITION
BF502026 810 bp mRNA linear EST 19-APR-2001
AT17637.5prime AT Drosophila melanogaster adult testes P0877
Drosophila melanogaster cDNA clone AT17637 5 similar to Octstt3:
FBan0007748 'enzyme' located on: 3R 96S16-96B17: 04/09/2001, mRNA
sequence.
ACCESSION
BF502026
VERSION
BF502026.2 GI:13695997
KEYWORDS
EST
Drosophila melanogaster (fruit fly)
SOURCE
Drosophila melanogaster
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Nephrotoidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 810)
AUTHORS
Stapleton,M., Brockstein,P., Hong,L., Agbayani,A., Baxter,E.,
Berman,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V.,
Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N.,
Li,P., Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunco,J.,
Paclet,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C.,
Lewis,S.E., Celniker,S. and Rubin,G.M.
BDGP/HMI AT Drosophila EST Project
Unpublished (2000)
On Dec 6, 2000 this sequence version replaced gi:11585327.
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003569; arm:X [20544566.20835979]
estimated-cyto:19E4-19F1: 04/09/2001 hit P element 1(3)j2D9:
1(3)j2D9 AQ026308 inserted at base 292 5' end of P element Inverse
PCR: 02/10/2001
Plate: AT.176 row: D column: 1
High quality sequence stop: 713.
Location/Qualifiers
1. 810
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="AT17637"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
AT.121-AT.319: DHS-alpha TonA"

FEATURES

source
Location/Qualifiers
1. 810
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="AT17637"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
AT.121-AT.319: DHS-alpha TonA"

/clone lib="Ar Drosophila melanogaster adult testes potB7"
/note="Organ: ADULT testes; Vector: potB7; Site 1: EcoRI;
Site 2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into potB7. Plasmid cDNA library."

ORIGIN

Query Match	30.4%;	Score 734;	DB 2;	Length 810;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches	784;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;

Qy	23	GCATTTCAGATCGGTTTAAATTTTCGAGTTACTGGCTGGAATTTGGACATGAATCGGACG	82
Db	1	GCATTTCAGATCGGTTTAAATTTTCGAGTTACTGGCTGGAATTTGGACATGAATCGGACG	60
Qy	83	CCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGAGCTTAATCACTTCGCCATCTCTG	142
Db	61	CCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGAGCTTAATCACTTCGCCATCTCTG	120
Qy	143	CTAATCGGCTGGCTGGCGGGAATTTCTCTCGGCTCTTTCGGCTGATCGGTTTCGAGTCG	202
Db	121	CTAATCGGCTGGCTGGCGGGAATTTCTCTCGGCTCTTTCGGCTGATCGGTTTCGAGTCG	180
Qy	203	ATTATCCATGATGTTTGAATCGGTTCAACTACCGGGCCACCGCTTACATGTTGAGAAAT	262
Db	181	ATTATCCATGATGTTTGAATCGGTTCAACTACCGGGCCACCGCTTACATGTTGAGAAAT	240
Qy	263	GGTTGGTACAACTTCCTCAACTGTTTTCGAGAGCGGCGATGATCGGTCGGTCGAGGATT	322
Db	241	GGTTGGTACAACTTCCTCAACTGTTTTCGAGAGCGGCGATGATCGGTCGGTCGAGGATT	300
Qy	323	GTGGGCGGTACCGTCTATCCCGGCTCATGATACGTCGGCGGAATCCATTGGCTGTG	382
Db	301	GTGGGCGGTACCGTCTATCCCGGCTCATGATACGTCGGCGGAATCCATTGGCTGTG	360
Qy	383	CACGTACTCAACATACCGGTCCTAATTCGTGACATCTGGTGTTCTGGCGCGATCTTC	442
Db	361	CACGTACTCAACATACCGGTCCTAATTCGTGACATCTGGTGTTCTGGCGCGATCTTC	420
Qy	443	AGTGGCCTGACCTCATCTCCACCTACTCTGTCACCAAGGAGTGTGGTCCGGCGGCC	502
Db	421	AGTGGCCTGACCTCATCTCCACCTACTCTGTCACCAAGGAGTGTGGTCCGGCGGCC	480
Qy	503	GCCCTCTTCGGCGGCGAGTTCATCGGCATCTGGCTGGCTGATCATGATGATGCTGGCT	562
Db	481	GCCCTCTTCGGCGGCGAGTTCATCGGCATCTGGCTGGCTGATCATGATGATGCTGGCT	540
Qy	563	GGATCGTACGATAACGAGGCGATTCGCCATATTCGCCCTGCGATTCACCTTCTCTGTG	622
Db	541	GGATCGTACGATAACGAGGCGATTCGCCATATTCGCCCTGCGATTCACCTTCTCTGTG	600
Qy	623	GTGGCTCAGTGAAGACATGATCCGTTCTGGTGGCGCGCGCCCTTTGCTTACTTTC	682
Db	601	GTGGCTCAGTGAAGACATGATCCGTTCTGGTGGCGCGCGCCCTTTGCTTACTTTC	660
Qy	683	TACATGGTGTCCGCTGGGTGGCTAGCTGTTCATCATCAACCTGATACCCCTTGCACGTC	742
Db	661	TACATGGTGTCCGCTGGGTGGCTAGCTGTTCATCATCAACCTGATACCCCTTGCACGTC	720
Qy	743	TTGCTACTGCTCATATGGGCGAGTACTCGCGGCTGTCTGACGACTACAGCACTTC	802
Db	721	TTGCTACTGCTCATATGGGCGAGTACTCGCGGCTGTCTGACGACTACAGCACTTC	780
Qy	803	TACAT 807	
Db	781	TACAT 785	

RESULT 2
CK659033
LOCUS

681 bp mRNA linear EST 30-JAN-2004

DEFINITION LP16462.Sprime LP Drosophila melanogaster larval-early pupal pot2
Drosophila melanogaster cDNA clone LP16462 5, mRNA sequence.
ACCESSION CK659033
VERSION CK659033.1 GI:41402558
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 681)
AUTHORS Harvey D., Brokstein P., Hong L., Evans-Holm M., Su C., Tsang G.,
Lewis S., and Rubin G.M.
BDGP/HMIR Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: LP164 row: F column: 2
High quality sequence stop: 670.
FEATURES
1..681
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="LP16462"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DHS-alpha"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pot2"
/note="Organ: whole body; Vector: pot2; Site 1: EcoRI;
Site 2: XhoI; Sized fractionated cDNAs were directly
ligated into pot2. Plasmid cDNA library."

ORIGIN

Query Match	28.2%;	Score 681;	DB 7;	Length 681;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	681;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;

Qy	16	TGTCGTTCATTCAGATCGGTTATAATTTTCGAGTTACTGGCTGGAATTTGGACATGAA	75
Db	1	TGTCGTTCATTCAGATCGGTTATAATTTTCGAGTTACTGGCTGGAATTTGGACATGAA	60
Qy	76	TCGGACGCCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGAGCCTAATACCTTCGC	135
Db	61	TCGGACGCCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGAGCCTAATACCTTCGC	120
Qy	136	CATCTCTGCTAAATCGGCTGGCGGATTTTCCTCTCGCCTTTCGCGCTCATCCGTTT	195
Db	121	CATCTCTGCTAAATCGGCTGGCGGATTTTCCTCTCGCCTTTCGCGCTCATCCGTTT	180
Qy	196	CGAGTCGATTTATCCATGATTTGATCCGTTGTTCAACTACCGGCCACCGCTACATGTT	255
Db	181	CGAGTCGATTTATCCATGATTTGATCCGTTGTTCAACTACCGGCCACCGCTACATGTT	240
Qy	256	GCAGATGTTGGTTCACACTTCTCACTGGTTGACGAGCGCGCATGTTATCCGCTCGG	315
Db	241	GCAGATGTTGGTTCACACTTCTCACTGGTTGACGAGCGCGCATGTTATCCGCTCGG	300
Qy	316	CAGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGCGGAATCCATTG	375
Db	301	CAGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGCGGAATCCATTG	360
Qy	376	GCTGCTGACGACTCAACATACCGGTCCATATTCGTGACATCTGCTGTTCTTGGGCGC	435
Db	361	GCTGCTGACGACTCAACATACCGGTCCATATTCGTGACATCTGCTGTTCTTGGGCGC	420
Qy	436	GATCTTCAGTGGCGGTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGTGTCGGC	495

Db 421 GAATTCAGTGGCCCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGGTCGC 480
 QY 496 GGGCGCGGCTCTTTGGCGGAGCTTTCATGCCATCGTGGCTGGCTACATCAGTAGTGC 555
 Db 481 GGGCGCGGCTCTTTGGCGGAGCTTTCATGCCATCGTGGCTGGCTACATCAGTAGTGC 540
 QY 556 GTGGCTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 615
 Db 541 GTGGCTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 QY 616 CTTGGTGGTGGCTCAGTGAAGACTGGATCCGCTGTTCTGGTGGCGGAGCGCTTTGTC 675
 Db 601 CTTGGTGGTGGCTCAGTGAAGACTGGATCCGCTGTTCTGGTGGCGGAGCGCTTTGTC 660
 QY 676 CTACTTCTACATGGTGTCCGC 696
 Db 661 CTACTTCTACATGGTGTCCGC 681

RESULT 3
 BF503420 701 bp mRNA linear EST 19-APR-2001
 LOCUS AT19406.5prime AT Drosophila melanogaster adult testes pOTB7
 DEFINITION Drosophila melanogaster cDNA clone AT19406 5 similar to OstStt3:
 FBan0007748 'enzyme' located on: 3R 96B16-96B17; : 04/09/2001, mRNA
 sequence.

ACCESSION BF503420
 VERSION BF503420.2 GI:13697340
 KEYWORDS EST.

SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 701)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E.,
 Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V.,
 Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N.,
 Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunco, J.,
 Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,
 Lewis, S.E., Celnik, S. and Rubin, G.M.

TITLE BDGP/HMI AT Drosophila EST Project
 JOURNAL Unpublished (2000)
 COMMENT On Dec 6, 2000 this sequence version replaced gi:11586721.

BDGP
 Contact: Stapleton, M.
 Lawrence Berkeley National Lab
 One Cyclotron Rd. Berkeley, CA 94720, USA
 Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit genomic AE003750: arm:3R [20671983,20899380]
 estimated-cyto:96B10-96C1: 04/09/2001
 Plate: AT.194 row: A column: 6
 High quality sequence stop: 602.

Location/Qualifiers
 1..701
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="AT19406"
 /sex="male"
 /dev_stage="0-3 day old Ore-R males"
 /lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
 AT.121-AT.319: DHS-alpha Tona"

/clone_lib="Ar Drosophila melanogaster adult testes pOTB7"
 /note="Organ: ADULT testes; Vector: pOTB7; Site 1: EcoRI;
 Site 2: XhoI; The mRNA for the testis library was made
 from testes and seminal vesicles hand dissected from 0-3
 day old Ore-R males. RNA kindly provided by the lab of
 Margaret Fuller. Sized fractionated cDNAs were directly
 ligated into pOTB7. Plasmid cDNA library."

ORIGIN

Query Match 27.5%; Score 664; DB 2; Length 701;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 CGTCTTCGTAAGTCTCATTAATGGGCAAGTACTCGCGCGTCTGCTGACAGTACAGCAC 798
 Db 1 CGTCTTCGTAAGTCTCATTAATGGGCAAGTACTCGCGCGTCTGCTGACAGTACAGCAC 60

QY 799 CTTCTACATCTCTGGGACTGCTGTTCTCCATGAGATCCCTTCGTTGGGATTCACACCGAT 858
 Db 61 CTTCTACATCTCTGGGACTGCTGTTCTCCATGAGATCCCTTCGTTGGGATTCACACCGAT 120

QY 859 ACGCACAGTGAACATCGGCTGGAGTGTGCTGCTCTTATGCGCGTGGCCAC 918
 Db 121 ACGCACAGTGAACATCGGCTGGAGTGTGCTGCTCTTATGCGCGTGGCCAC 180

QY 919 CTTGGCGCATTTGCACTCGTGTGCGCAACGAGTCCGGAAGCTGTTCAATGCTCGG 978
 Db 181 CTTGGCGCATTTGCACTCGTGTGCGCAACGAGTCCGGAAGCTGTTCAATGCTCGG 240

QY 979 CGGATTCGTTGGGCGTTGGCGTCTTGTGGCGCTGCTGCTGCTCACCATCTGGGCT 1038
 Db 241 CGGATTCGTTGGGCGTTGGCGTCTTGTGGCGCTGCTGCTGCTCACCATCTGGGCT 300

QY 1039 TGTGGCGCGTGGAGTGGAGCGTTCTACTCGCTGTGGATAGTGGCTAGCCAAAGATCA 1098
 Db 301 TGTGGCGCGTGGAGTGGAGCGTTCTACTCGCTGTGGATAGTGGCTAGCCAAAGATCA 360

QY 1099 CATTCCCATCATTCGATCCGTCGAGCATCAGCCACCATCTGCTGCTGCTGCTTCTT 1158
 Db 361 CATTCCCATCATTCGATCCGTCGAGCATCAGCCACCATCTGCTGCTGCTGCTTCTT 420

QY 1159 TGATCTGCACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1218
 Db 421 TGATCTGCACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480

QY 1219 CAACACAGAGCGCGTTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1278
 Db 481 CAACACAGAGCGCGTTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

QY 1279 GATGTGGCTTTGATGTTGACCCCTCACGCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1338
 Db 541 GATGTGGCTTTGATGTTGACCCCTCACGCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 600

QY 1339 TTCCGGGACTGTTGGATGTTCTTCGACAGAGATTCGTTAAGCGAATGGCAGACCAT 1398
 Db 601 TTCCGGGACTGTTGGATGTTCTTCGACAGAGATTCGTTAAGCGAATGGCAGACCAT 660

QY 1399 AAGC 1402
 Db 661 AAGC 664

RESULT 4
 BF503420
 LOCUS

DEFINITION RE43425.5prime RE Drosophila melanogaster normalized Embryo pFic-1
 Drosophila melanogaster cDNA clone RE43425 5 similar to OstStt3:
 FBan0007748 'enzyme' located on: 3R 96B16-96B17; : 05/13/2001, mRNA
 sequence.

ACCESSION BF503420
 VERSION BF503420.1 GI:15051528
 KEYWORDS EST.

SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 666)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,

666 bp mRNA linear EST 31-JUL-2001
 RE43425.5prime RE Drosophila melanogaster normalized Embryo pFic-1
 Drosophila melanogaster cDNA clone RE43425 5 similar to OstStt3:
 FBan0007748 'enzyme' located on: 3R 96B16-96B17; : 05/13/2001, mRNA
 sequence.

BF503420
 BF503420.1 GI:15051528
 EST.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 666)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,

666 bp mRNA linear EST 31-JUL-2001
 RE43425.5prime RE Drosophila melanogaster normalized Embryo pFic-1
 Drosophila melanogaster cDNA clone RE43425 5 similar to OstStt3:
 FBan0007748 'enzyme' located on: 3R 96B16-96B17; : 05/13/2001, mRNA
 sequence.

BF503420
 BF503420.1 GI:15051528
 EST.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 666)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,

666 bp mRNA linear EST 31-JUL-2001
 RE43425.5prime RE Drosophila melanogaster normalized Embryo pFic-1
 Drosophila melanogaster cDNA clone RE43425 5 similar to OstStt3:
 FBan0007748 'enzyme' located on: 3R 96B16-96B17; : 05/13/2001, mRNA
 sequence.

BF503420
 BF503420.1 GI:15051528
 EST.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 666)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,

666 bp mRNA linear EST 31-JUL-2001
 RE43425.5prime RE Drosophila melanogaster normalized Embryo pFic-1
 Drosophila melanogaster cDNA clone RE43425 5 similar to OstStt3:
 FBan0007748 'enzyme' located on: 3R 96B16-96B17; : 05/13/2001, mRNA
 sequence.

BF503420
 BF503420.1 GI:15051528
 EST.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 666)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,

QY	1	TCTAAGCGAAGATGTGTCGTTGCGATTTCAGATCGGTTATAATTTTCAGATTACGCGTG	60
Db	9	TCTAAGCGAAGATGTGTCGTTGCGATTTCAGATCGGTTATAATTTTCAGATTACGCGTG	68
QY	61	GAAATTGGGACATGAATCGGACGCGGAAGATGCTGAACAGACAAGTGGTGGCTACACAG	120
Db	69	GAAATTGGGACATGAATCGGACGCGGAAGATGCTGAACAGACAAGTGGTGGCTACACAG	128
QY	121	CCTAATACCTTTCGCCATTCGTGTAATCGGCTGGCTGGCCGGATTTTCTCTCGGCCTTT	180
Db	129	CCTAATACCTTTCGCCATTCGTGTAATCGGCTGGCTGGCCGGATTTTCTCTCGGCCTTT	188
QY	181	CGCGCTCATCGTTTCAGATCGATTATTCATGATGTTTGATCCGTGGTTCAACTACCGGCG	240
Db	189	CGCGCTCATCGTTTCAGATCGATTATTCATGATGTTTGATCCGTGGTTCAACTACCGGCG	248
QY	241	CACCGCTACATGTCGAGAAATGGTTGGTAAACCTTCCTCAACTGGTTTCGACGACGCGCG	300
Db	249	CACCGCTACATGTCGAGAAATGGTTGGTAAACCTTCCTCAACTGGTTTCGACGACGCGCG	308
QY	301	ATGGTATCCGCTCGCGAGGATGTGGCGGTACCGTCTATCCGGCGCTGATGATTACGTC	360
Db	309	ATGGTATCCGCTCGCGAGGATGTGGCGGTACCGTCTATCCGGCGCTGATGATTACGTC	368
QY	361	CGGCGGAATCCATTGGCTGTGCACTGATCAACATACCGGTCCTATTCGTCGATCTG	420
Db	369	CGGCGGAATCCATTGGCTGTGCACTGATCAACATACCGGTCCTATTCGTCGATCTG	428
QY	421	CGTGTTCCTGGCGCGATCTTCAGTGGCTCGACTCCATCTCCACCTACTGCTCAACAA	480
Db	429	CGTGTTCCTGGCGCGATCTTCAGTGGCTCGACTCCATCTCCACCTACTGCTCAACAA	488
QY	481	GGAGCTGTGGTCCGCGGCGCGCGCTCTTCGCGCGCAGGCTTCATCGCCATCTGTCGCTGG	540
Db	489	GGAGCTGTGGTCCGCGGCGCGCGCTCTTCGCGCGCAGGCTTCATCGCCATCTGTCGCTGG	548

QY

61 GAATTGGGACATGAATCGGACGCCGAAGATGC

1 TCTACGCGAAGAAATGTCGTTGCATCTTCAGATCGGTTATAATTTTCGAGTTACTGGCTG 60
11 TCTACGCGAAGAAATGTCGTTGCATCTTCAGATCGGTTATAATTTTCGAGTTACTGGCTG 70
61 GAATTGGGACATGAATTCGGACGCCGAAGATGTTGAACAGCAGGTGGCTGGCTACAGCAG 120

```

Db 71 GAATGGGACATGAATCGGACGCGGAGATGCTGAACAGCAAGGTGGCTGGCTACAGAG 130
QY 121 CTAATACACTTCGCCATCTCTGTAATCGCTGGCTGGCGGAGATTTTCCTCTGCTCTTT 180
Db 131 CTAATACACTTCGCCATCTCTGTAATCGCTGGCTGGCGGAGATTTTCCTCTGCTCTTT 190
QY 181 CGCGGTCATCCGTTTCGAGTGGATTCATCGATGTTGATCCGTTCACTACCGGC 240
Db 191 CGCGGTCATCCGTTTCGAGTGGATTCATCGATGTTGATCCGTTCACTACCGGC 250
QY 241 CACCGCTACATGTCGAGATGTTGTTGATCAACTTCTCAACTGTTGACGAGCGGC 300
Db 251 CACCGCTACATGTCGAGATGTTGTTGATCAACTTCTCAACTGTTGACGAGCGGC 310
QY 301 ATGGTATCCGTCGCGAGATGTTGGCGGTCACGTTATCCCGGCTGATGATACGTC 360
Db 311 ATGGTATCCGTCGCGAGATGTTGGCGGTCACGTTATCCCGGCTGATGATACGTC 370
QY 361 CGCGGATTCATGTCGAGTTCAGTACTCAACATACCGTCCATATTCGTGACATCTG 420
Db 371 CGCGGATTCATGTCGAGTTCAGTACTCAACATACCGTCCATATTCGTGACATCTG 430
QY 421 CTTGTTCTCGGCGGATTCAGTGGCTGACCTCCATCTCCACCTACCTGTCACCA 480
Db 431 CTTGTTCTCGGCGGATTCAGTGGCTGACCTCCATCTCCACCTACCTGTCACCA 490
QY 481 GAGTGTGTCGCGGCGGCTCTTCGCGGAGTTCATCCGCTGTCGCTG 540
Db 491 GAGTGTGTCGCGGCGGCTCTTCGCGGAGTTCATCCGCTGTCGCTG 550
QY 541 CTACATGATGTCGCGGCTGATGTCGATGATGATGATGATGATGATGATGATGATG 600
Db 551 CTACATGATGTCGCGGCTGATGTCGATGATGATGATGATGATGATGATGATGATG 610
QY 601 GCAGTTCACCTACTTCTCTGGTGGTGGCTCAGTGAAGACTGGATCCGTTCTGTCG 658
Db 611 GCAGTTCACCTACTTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 668

```

```

RESULT 6
CK657451
LOCUS
DEFINITION
  LP23759: Spine LP Drosophila melanogaster larval-early pupal P0T2
  Drosophila melanogaster cDNA clone LP23759 5, mRNA sequence.
ACCESSION
  CK657451
VERSION
  1
KEYWORDS
  EST.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 715)
  Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
  Lewis, S. and Rubin, G.M.
  BDGP/HMI Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST\_est@fruitfly.berkeley.edu
  Plate: LP.237 row: E column: 11
  High quality sequence, scop: 644.
  Location/Qualifiers
    1..715
      /organism="Drosophila melanogaster"
      /mol_type="mRNA"
      /db_xref="taxon:7227"
      /clone="LP23759"
      /sex="male and female"
      /dev_stage="larvae-pupae"

```

FEATURES

source

```

ACCESSION
  B1171940
VERSION
  B1171940
KEYWORDS
  EST.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.

```

```

/lab host="DHS-alpha"
/clone_lib="LP Drosophila melanogaster larval-early pupal
P0T2"
/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI;
Site 2: XbaI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."

```

```

ORIGIN
Query Match      27.2% Score 657; DB 7; Length 715;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 657; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GGCTGGATTTGGGACATGATCGGACGCGGAGATGCTGAACAGCAAGGTGGCTGGCTAC 115
Db 45 GGCTGGATTTGGGACATGATCGGACGCGGAGATGCTGAACAGCAAGGTGGCTGGCTAC 104
QY 116 AGCAGCTTAATCACTTTGCCATCTCTGCTAATCACTGCTGGCTGGCGGATTTTCCTCTCG 175
Db 105 AGCAGCTTAATCACTTTGCCATCTCTGCTAATCACTGCTGGCTGGCGGATTTTCCTCTCG 164
QY 176 CTCTTCGCGTCATCCGTTTCGAGTGGATTCATCCATGATTTGATCCGTTGTTCACTAC 235
Db 165 CTCTTCGCGTCATCCGTTTCGAGTGGATTCATCCATGATTTGATCCGTTGTTCACTAC 224
QY 236 CGGCGCAACCGCTACATGTTGTCAGAAATGTTGGTACAACTTCCTCAACTGGTTGCAAG 295
Db 225 CGGCGCAACCGCTACATGTTGTCAGAAATGTTGGTACAACTTCCTCAACTGGTTGCAAG 284
QY 296 CGGCGCAACCGCTACATGTTGTCAGAAATGTTGGTACAACTTCCTCAACTGGTTGCAAG 355
Db 285 CGGCGCAACCGCTACATGTTGTCAGAAATGTTGGTACAACTTCCTCAACTGGTTGCAAG 344
QY 356 ACCTTCGCGGGAATPCCATTTGGTCTGCTGACGTAATCAACATACCGTTCATATTCGTGAC 415
Db 345 ACCTTCGCGGGAATPCCATTTGGTCTGCTGACGTAATCAACATACCGTTCATATTCGTGAC 404
QY 416 ATCTTCGCTGTTCCGCGGCGGATCTTCAGTGGCTGACGTCATCCATCCACCTACCTGCTG 475
Db 405 ATCTTCGCTGTTCCGCGGCGGATCTTCAGTGGCTGACGTCATCCATCCACCTACCTGCTG 464
QY 476 ACCAAGAGCTGTGGTTCGCGGCGGCGGCTCTTCGCGGCGGAGTTCATCCGCTGCTG 535
Db 465 ACCAAGAGCTGTGGTTCGCGGCGGCGGCTCTTCGCGGCGGAGTTCATCCGCTGCTG 524
QY 536 CCTGCTACATCATGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 595
Db 525 CCTGCTACATCATGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 584
QY 596 GCGCTGAGTTTCACTACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 655
Db 585 GCGCTGAGTTTCACTACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 644
QY 656 TCGGCGGCGGCGGCTTGTCTCTACTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 712
Db 645 TCGGCGGCGGCGGCTTGTCTCTACTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 701

```

```

RESULT 7
B1171940
LOCUS
DEFINITION
  RE13841: Spine RE Drosophila melanogaster normalized Embryo pPle-1
  Drosophila melanogaster cDNA clone RE13841.5 similar to OctStc3:
  FBan0007748 'enzyme' located on: 3R 96B16-96B17; 04/06/2001, mRNA
  sequence.
  B1171940
  B1171940.1
  GI:14637747
  EST.
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.

```

```

REFERENCE
AUTHORS
  1 (bases 1 to 652)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Chavez,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  hit genomic AE003750: arm:3R [20671983,20899380]
  estimated-cyto:96B10-96C1: 04/06/2001 hit P element 1(3)j2D9:
  1(3)j2D9 AQ026308 inserted at base 292 5' end of P element Inverse
  PCR: 04/06/2001
  Plate: RE.138 row: D column: 5
  High quality sequence stop: 590.
  Location/Qualifiers
    1..652
    /organism="Drosophila melanogaster"
    /mol_type="mRNA"
    /db_xref="taxon:7227"
    /sex="male and female"
    /dev_stage="0-24 hours mixed stage embryonic"
    /lab_hosts="DHS-alpha TorA"
    /clone_lib="RE Drosophila melanogaster normalized Embryo
    pf1c-1"
    /note="Organ: embryo; Vector: pELc1; Site: 1: XhoI; Site: 2:
    BamHI; Library was kindly generated by Piero Carninci at
    the RIKEN. The library was normalized and excised using
    Cre recombinase. Plasmid cDNA library."

FEATURES
source
  1..652
  /organism="Drosophila melanogaster"
  /mol_type="mRNA"
  /db_xref="taxon:7227"
  /sex="male and female"
  /dev_stage="0-24 hours mixed stage embryonic"
  /lab_hosts="DHS-alpha TorA"
  /clone_lib="RE Drosophila melanogaster normalized Embryo
  pf1c-1"
  /note="Organ: embryo; Vector: pELc1; Site: 1: XhoI; Site: 2:
  BamHI; Library was kindly generated by Piero Carninci at
  the RIKEN. The library was normalized and excised using
  Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match      27.0%; Score 652; DB 4; Length 652;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AATGTCGTCGATTCAGATCGGTTAATTTTCAGTTACTGCTGGATTCGGACA 71
DB 1 AATGTCGTCGATTCAGATCGGTTAATTTTCAGTTACTGCTGGATTCGGACA 60

QY 72 TGAATCGGACCCGGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAGCTTAATCACT 131
DB 61 TGAATCGGACCCGGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAGCTTAATCACT 120

QY 132 TCGGCATCCCTGCTAATCGCTGCTGCTGGCGGATTTTCCTCTCGCTCTTCGCGTCATCC 191
DB 121 TCGGCATCCCTGCTAATCGCTGCTGCTGGCGGATTTTCCTCTCGCTCTTCGCGTCATCC 180

QY 192 GTTTCGAGTCGATTATCCATGAGTTTGATCCGCTGGTTCAACTACCGGGCACCGCTACA 251
DB 181 GTTTCGAGTCGATTATCCATGAGTTTGATCCGCTGGTTCAACTACCGGGCACCGCTACA 240

QY 252 TGGTGCAGATGGTTGGTGAACAATCTCTCACTGGTTTCGACGAGCGCGATGTTATCCGC 311
DB 241 TGGTGCAGATGGTTGGTGAACAATCTCTCACTGGTTTCGACGAGCGCGATGTTATCCGC 300

QY 312 TCGGACGAGATTGTGGCGGACCGTCTATCCCGGCTGATGATTACGTCGGCGGAATCC 371
DB 301 TCGGACGAGATTGTGGCGGACCGTCTATCCCGGCTGATGATTACGTCGGCGGAATCC 360

QY 372 ATTGGCTGCTCAGTACTCAACATACCGGTCCATATTCGTGACATCTGCGTGTTCCTGG 431
DB 361 ATTGGCTGCTCAGTACTCAACATACCGGTCCATATTCGTGACATCTGCGTGTTCCTGG 420

QY 432 CGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGCTACCAAGGAGCTGGGT 491
DB 421 CGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGCTACCAAGGAGCTGGGT 480

```

```

QY 492 CGCGGGGGCGGGCGCTCTTCGCGGGCAGCTTCATCGCCATCGTGGCTGCTACATCAGTA 551
DB 481 CGCGGGGGCGGGCGCTCTTCGCGGGCAGCTTCATCGCCATCGTGGCTGCTACATCAGTA 540
QY 552 GGTCTGGTGGCTGGATCGTACGATTAACAGAGGCAATTCGCATATTCGCGCTGCAGTTCACT 611
DB 541 GGTCTGGTGGCTGGATCGTACGATTAACAGAGGCAATTCGCATATTCGCGCTGCAGTTCACT 600
QY 612 ACTTCTCTGGTGGCTGGCTCAGTGAAGACTGGATCGTGGTCTGGTGGCGGC 663
DB 601 ACTTCTCTGGTGGCTGGCTCAGTGAAGACTGGATCGTGGTCTGGTGGCGGC 652

RESULT 8
LOCUS      BG636414
DEFINITION SD14123: Sprime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD14123 5 similar to
            OstBt3: FBan0007748 'enzyme' located on: 3R 96B16-96B17;:
            04/13/2001, mRNA sequence.
ACCESSION  BG636414
VERSION    BG636414
KEYWORDS   EST.
SOURCE     BG636414.1 GI:13763951
ORGANISM   Drosophila melanogaster (fruit fly)
            Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 644)
AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Teang,G.,
            Lewis,S. and Rubin,G.M.
TITLE     BDGP/HMI Drosophila EST Project
JOURNAL   Unpublished (2001)
COMMENT   Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003750: arm:3R [20671983,20899380]
            estimated-cyto:96B10-96C1: 04/13/2001 hit P element 1(3)j2D9:
            1(3)j2D9 AQ026308 inserted at base 292 5' end of P element Inverse
            PCR: 03/20/2001
            Plate: SD.141 row: B column: 11
            High quality sequence stop: 612.
            Location/Qualifiers
              1..644
              /organism="Drosophila melanogaster"
              /mol_type="mRNA"
              /db_xref="taxon:7227"
              /clone="SD14123"
              /lab_hosts="DHS-alpha"
              /clone_lib="SD Drosophila melanogaster Schneider L2 cell
              culture pOT2"
              /note="Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized
              fractionated cDNAs were directly ligated into pOT2.
              Plasmid cDNA library."

ORIGIN
Query Match      26.6%; Score 644; DB 4; Length 644;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAAGCAGGAATGTGCTGCTTCAGATTCGATTCGATTCGATTCGATTCGATTCGCTG 60
DB 1 TCTAAGCAGGAATGTGCTGCTTCGATTCGATTCGATTCGATTCGATTCGATTCGCTG 60

QY 61 GAATTCGACATGATTCGACCGCGAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120
DB 61 GAATTCGACATGATTCGACCGCGAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120
QY 121 CCTAATACCTTCGCTCAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 180

```



```

Db      121  CCTAATACCTTCGCGCATCTCTGTAATCGCTGGCTGCGCGAATTTTCCTCTCGCCTTT 180
QY      181  GCGCGTCATCGTTTCGATGATTCATGATGATTCATGATGATTCATGATGATTCATGATGAT 240
Db      181  GCGCGTCATCGTTTCGATGATTCATGATGATTCATGATGATTCATGATGATTCATGATGAT 240
QY      241  CACCGCCTACATGTTGCGAGATGTTGTTGATGATTCATGATGATTCATGATGATTCATGATGAT 300
Db      241  CACCGCCTACATGTTGCGAGATGTTGTTGATGATTCATGATGATTCATGATGATTCATGATGAT 300
QY      301  ATGGTATCGCTCGCGAGATGTTGCGAGATGTTGCGAGATGTTGCGAGATGTTGCGAGATGTTG 360
Db      301  ATGGTATCGCTCGCGAGATGTTGCGAGATGTTGCGAGATGTTGCGAGATGTTGCGAGATGTTG 360
QY      361  CGCGGGAATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db      361  CGCGGGAATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY      421  CGTGTTCCTGCGCGAGATTCCTGATGATTCATGATGATTCATGATGATTCATGATGATTCATG 480
Db      421  CGTGTTCCTGCGCGAGATTCCTGATGATTCATGATGATTCATGATGATTCATGATGATTCATG 480
QY      481  GGAGCTGTGCTCGCGCGCGCGCGCTCTTCCGCGCGCGCGCTCTTCCGCGCGCGCGCTCTTCC 540
Db      481  GGAGCTGTGCTCGCGCGCGCGCGCTCTTCCGCGCGCGCGCTCTTCCGCGCGCGCGCTCTTCC 540
QY      541  CTACATCAGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db      541  CTACATCAGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY      601  GCAGTTCACCTACCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 644
Db      601  GCAGTTCACCTACCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 644

```

RESULT 9
BI374334

LOCUS
DEFINITION

BI374334 654 bp mRNA linear EST 01-AUG-2001
RES1893: Spine RE Drosophila melanogaster normalized Embryo pf1c-1
Drosophila melanogaster cDNA clone RE61893 5 similar to OatStt3;
F800007748 GO: [oligoaccharide transferase (GO:0004576); enzyme
(GO:0003824)] located on: 3R 96B16-96B17; 05/16/2001, mRNA
sequence.

ACCESSION
VERSION

KEYWORDS
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BDGP/HMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
hit genomic AB003750: arm3R [20671983, 20899380]
estimated-cyto:96B10-96C1: 05/16/2001 hit P element 1(3)j2D9:
1(3)j2D9 AQ026308 inserted at base 292 5', end of P element Inverse
PCR: 05/16/2001
Plate: RE.618 row: H column: 9
High quality sequence stop: 595.

FEATURES
source

Location/Qualifiers
1..654
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE61893"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pf1c-1"
/notes="Organ: embryo; Vector: pf1c1; Site 1: XhoI; Site 2:
BamHI; library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN

```

Query Match      26.6%; Score 644; DB 4; Length 654;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TCTAAGCGAAGATGTCGTTGCAATTCAGATCGGTATATATTTTCGAGTTACTGGCTG 60
Db      11  TCTAAGCGAAGATGTCGTTGCAATTCAGATCGGTATATATTTTCGAGTTACTGGCTG 70
QY      61  GAATTGGGACATGAATCGGACCCGGAAGATGCTGAACAGCAAGTGGTGGCTACAGCAG 120
Db      71  GAATTGGGACATGAATCGGACCCGGAAGATGCTGAACAGCAAGTGGTGGCTACAGCAG 130
QY      121  CTTAATCACCTTCGCGATCCTCTAATCGCCTGGCTGGCGGATTTCTCTCGGCTCTT 180
Db      131  CTTAATCACCTTCGCGATCCTCTAATCGCCTGGCTGGCGGATTTCTCTCGGCTCTT 190
QY      181  CGCGGTTCATCCGTTTCGAGTTCGATTCATCGATGTTGATCGGTGGTTCACATACCGGCG 240
Db      191  CGCGGTTCATCCGTTTCGAGTTCGATTCATCGATGTTGATCGGTGGTTCACATACCGGCG 250
QY      241  CACCGCTACATGGTGCAGATGGTGGTGAACACTTCCTCACTGGTTCGAGAGCGGCG 300
Db      251  CACCGCTACATGGTGCAGATGGTGGTGAACACTTCCTCACTGGTTCGAGAGCGGCG 310
QY      301  ATGGTATCGCTCGCGAGATTTGGGGGTACCGTCTATCCGCCCTGATGATTACGTC 360
Db      311  ATGGTATCGCTCGCGAGATTTGGGGGTACCGTCTATCCGCCCTGATGATTACGTC 370
QY      361  CGCGGGAATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db      371  CGCGGGAATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 430
QY      421  CGTGTTCCTGCGCGAGATTCCTGATGATTCATGATGATTCATGATGATTCATGATGATTCATG 480
Db      431  CGTGTTCCTGCGCGAGATTCCTGATGATTCATGATGATTCATGATGATTCATGATGATTCATG 490
QY      481  GGAGTTCACCTACCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db      491  GGAGTTCACCTACCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 550
QY      541  CTACATCAGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db      551  CTACATCAGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 610
QY      601  GCAGTTCACCTACCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 644
Db      611  GCAGTTCACCTACCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 654

```

RESULT 10
A1257750

LOCUS
DEFINITION

A1257750 641 bp mRNA linear EST 19-APR-2001
LP06212: Spine LP Drosophila melanogaster larval-early pupal pOT2
Drosophila melanogaster cDNA clone LP06212 5prime, mRNA sequence.
A1257750
A1257750.1 GI:3865275

1 (bases 1 to 642)

REFERENCE
AUTHORS
Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
Migra,S., Mungall,C.J., Nunco,J., Pacleb,J., Paragas,V., Park,S.,
Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Ceiniker,S. and
Rubin,G.M.

TITLE
JOURNAL
COMMENT
BDGP/HIMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AE003750: arm:3R [20671983, 20899380]
estimated-cyto:96B10-96C1: 04/12/2001 hit P element 1(3)J2D9:
1(3)J2D9 AQ026308 inserted at base 292 5' end of P element inverse
PCR: 03/19/2001
Plate: RE.157 row: G column: 2
High quality sequence stop: 518.
Location/Qualifiers
1. 642
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE15774"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha ToTA"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFLC-1"
/note="Organ: embryo; Vector: pFLC1; Site: 1: XhoI; Site: 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 26.3%; Score 635; DB 4; Length 642;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps: 0

QY 1 TCTAAGCGAAGATGTGTGCTTGATTTTCAGATCGGTTATAATTTTCGAGTTACTGGCTG 60
DB 1 TCTAAGCGAAGATGTGTGCTTGATTTTCAGATCGGTTATAATTTTCGAGTTACTGGCTG 67
QY 61 GAATTTGGGACATGAATCGGACGCCAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120
DB 68 GAATTTGGGACATGAATCGGACGCCAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 127
QY 121 CCTAATCACTTCGCGCATCTCTGAATCGCTGGCTGGCTGGCGGATTTTCCTTCGCGCTCTT 180
DB 128 CCTAATCACTTCGCGCATCTCTGAATCGCTGGCTGGCTGGCGGATTTTCCTTCGCGCTCTT 187
QY 181 CGCGCTCATCGTTTCGAGTCGATTATCCATGAGTTTGATTCGCTGGTTCAACTACCGGCG 240
DB 188 CGCGCTCATCGTTTCGAGTCGATTATCCATGAGTTTGATTCGCTGGTTCAACTACCGGCG 247
QY 241 CACCGCTTACATGGTCGAGATGTTGGTGAACAATTCCTCAACTGGTTTCGACAGCGGCG 300
DB 248 CACCGCTTACATGGTCGAGATGTTGGTGAACAATTCCTCAACTGGTTTCGACAGCGGCG 307
QY 301 ATGTTATCCGCTCGCGCAGGATTTGTTGGCGGTTACCGTCTATCCCGGCTCATGATTACGTC 360
DB 308 ATGTTATCCGCTCGCGCAGGATTTGTTGGCGGTTACCGTCTATCCCGGCTCATGATTACGTC 367
QY 361 CGCGGATTCATTTGGCTGTCGACGTACTCACATACCGGTTCCATTTCTGTGACATCTG 420
DB 368 CGCGGATTCATTTGGCTGTCGACGTACTCACATACCGGTTCCATTTCTGTGACATCTG 427
QY 421 CGTGTCTCGCGCGGATCTTCAGTGGCCCTGACCTCATCTCCACCTACCTGCTGACCAA 480
DB 428 CGTGTCTCGCGCGGATCTTCAGTGGCCCTGACCTCACTCACTCACTCACTCACTCACT 487

```
QY 189 TCCGTTTCAGTCGATTATCATCAGTTTGATCGTGGTTCAACTACCGGGCCACCGCCT 248
Db 181 TCCGTTTCAGTCGATTATCATCAGTTTGATCGTGGTTCAACTACCGGGCCACCGCCT 240
QY 249 ACATGGTGCAGATGGTGGTACAACTTCCTCAACTGGTTCGACGCGCGCATGTATC 308
Db 241 ACATGGTGCAGATGGTGGTACAACTTCCTCAACTGGTTCGACGCGCGCATGTATC 300
QY 309 CGCTCGCAGAGATTGTGGGGGTACCGCTATCCCGGCTGATGATTACGTCGGGGGAA 368
Db 301 CGCTCGCAGAGATTGTGGGGGTACCGCTATCCCGGCTGATGATTACGTCGGGGGAA 360
QY 369 TCCATTGGCTGCTCAGTACTCAACATACGGGTCCATATCGTGACATCGTGTTC 428
Db 361 TCCATTGGCTGCTCAGTACTCAACATACGGGTCCATATCGTGACATCGTGTTC 420
QY 429 TGGCGCCGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGACCAAGAGCTGT 488
Db 421 TGGCGCCGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGACCAAGAGCTGT 480
QY 489 GGTCCGGGGGGCGCGCTCTTCCGCGCAGCTTCATCGCCATCGTGGCTGATCA 548
Db 481 GGTCCGGGGGGCGCGCTCTTCCGCGCAGCTTCATCGCCATCGTGGCTGATCA 540
QY 549 GTAGTTCGGTGGCTGATCGTACGATACGAGGCGATTGCGCATATTCGCGCTCGAGTTCA 608
Db 541 GTAGTTCGGTGGCTGATCGTACGATACGAGGCGATTGCGCATATTCGCGCTCGAGTTCA 600
QY 609 CTTACTTCCTGTGGGTGGCTCA 631
Db 601 CTTACTTCCTGTGGGTGGCTCA 623
```

RESULT 15

BF496296

LOCUS

DEFINITION

BF496296 655 bp mRNA linear EST 19-APR-2001
Ari0060.5prime AT Drosophila melanogaster adult testes pOTB7
Drosophila melanogaster cDNA clone Ari0060 5 similar to OstStt3:
Fban007748 'enzyme' located on: 3R 96B16-96B17; 04/08/2001, mRNA
sequence.

ACCESSION

BF496296

VERSION

EST.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1. (bases 1 to 655)

AUTHORS

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E.,
Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V.,
Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N.,
Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C. J., Nunoo, J.,
Paclob, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,
Lewis, S. E., Celniker, S. and Rubin, G. M.
BDGP/HHMI AT Drosophila EST Project
Unpublished (2000)

JOURNAL

COMMENT

On Dec 6, 2000 this sequence version replaced gi:11579597.
Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit genomic AEO03750: arm:3R [20671983,20899380]

estimated-cyto:96B10-96C1: 04/08/2001 hit P element 1(3)j2D9:

1(3)j2D9 A0026308 inserted at base 292 5' end of P element Inverse

PCR: 02/09/2001

Plate: AT100 row: E column: 12

High quality sequence stop: 577.

Location/Qualifiers

1. .655

/organism="Drosophila melanogaster"

FEATURES

source

/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="AT10060"
/sex="male"

/dev_stage="0-3 day old Ore-R males"

/lab_host="Plates AT10-AT120: DHS-alpha. Plates

AT121-AT139: DHS-alpha Tona"

/clone_lib="AT Drosophila melanogaster adult testes pOTB7"

/note="Organ: ADULT testes; Vector: pOTB7; Site 1: EcORI;

Site 2: XhoI; The mRNA for the testis library was made

from testes and seminal vesicles hand dissected from 0-3

day old Ore-R males. RNA kindly provided by the lab of

Margaret Fuller. Sized fractionated cDNAs were directly

ligated into pOTB7. Plasmid cDNA library."

ORIGIN

```
Query Match 25.4%; Score 615; DB 2; Length 655;  
Best Local Similarity 100.0%; Pred. No. 2.3e-309; Indels 0; Gaps 0;  
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 58 CTGGAAATTGGGACATGAATCGGACCGCCGAAAGATGCTGAACAGCAAGGTGGTGGCTACAG 117  
Db 1 CTGGAAATTGGGACATGAATCGGACCGCCGAAAGATGCTGAACAGCAAGGTGGTGGCTACAG 60  
  
QY 118 CAGCCTAATACCTTCGCGCATCTCTGCTAATCGCCTGGCTGGCGGATTTCTCTCGCCT 177  
Db 61 CAGCCTAATACCTTCGCGCATCTCTGCTAATCGCCTGGCTGGCGGATTTCTCTCGCCT 120  
  
QY 178 CTTCCCGCTCATCCGTTTCGAGTCGATTCATCCATGAGTTTTCGCGTGGTTCAACTACCG 237  
Db 121 CTTCCCGCTCATCCGTTTCGAGTCGATTCATCCATGAGTTTTCGCGTGGTTCAACTACCG 180  
  
QY 238 GGCACACCGCTACATGGTGCAGAAATGGTGTGATCAACTTCCTCACTGGTTCGACGAGCG 297  
Db 181 GGCACACCGCTACATGGTGTGAGAAATGGTGTGATCAACTTCCTCACTGGTTCGACGAGCG 240  
  
QY 298 CGCATGGTATCCGCTCGGACGATTTGGGGCGGTACCGTCTATCCCGGCTGATGATTAC 357  
Db 241 CGCATGGTATCCGCTCGGACGATTTGGGGCGGTACCGTCTATCCCGGCTGATGATTAC 300  
  
QY 358 GTCCCGCGGAATCCATTCGCTGCTGCAGTACTCAACATACCGGTCCATATTCGTGACAT 417  
Db 301 GTCCCGCGGAATCCATTCGCTGCTGCAGTACTCAACATACCGGTCCATATTCGTGACAT 360  
  
QY 418 CTGCGTGTTCCTGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGAC 477  
Db 361 CTGCGTGTTCCTGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGAC 420  
  
QY 478 CAAGAGCTGTGGTCCGCGCGCGGCTCTTCGCGCCAGCTTCATTCGCCATTCGTGCC 537  
Db 421 CAAGAGCTGTGGTCCGCGCGCGGCTCTTCGCGCCAGCTTCATTCGCCATTCGTGCC 480  
  
QY 538 TGGCTACATCAGTAGTTCGCTGGTGGATCGTACGATTAACGAGGCGATTCGCATATTCGC 597  
Db 481 TGGCTACATCAGTAGTTCGCTGGTGGATCGTACGATTAACGAGGCGATTCGCATATTCGC 540  
  
QY 598 COTGAGTTCACCTACTTCCTGCTGGTGGCTGCTGAGTGAAGACTGGATTCGCTGCTGTC 657  
Db 541 COTGAGTTCACCTACTTCCTGCTGGTGGCTGCTGAGTGAAGACTGGATTCGCTGCTGTC 600  
  
QY 658 GGCOCGACCGCGTTT 672  
Db 601 GGCOCGACCGCGTTT 615
```

Search completed: December 16, 2004, 07:12:51

Job time : 11096.6 secs

Blank
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 22:52:10 ; Search time 10054.9 Seconds
(without alignments)
11367.538 Million cell updates/sec

Title: US-10-028-384-7
Perfect score: 2417
Sequence: 1 tctaaaggaggaagtgtctg.....ccaaaaaataaaaaa 2417

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053450

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ats.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2417	100.0	2417	3	AF132552
2	2417	100.0	2417	6	AX799088
3	2410.4	99.7	2699	6	CQ589353
C 4	2147.4	88.8	4922	6	CQ589352
C 5	2147.4	88.8	163921	3	AC007853
C 6	2147.4	88.8	181132	3	AC008206
C 7	2147.4	88.8	227219	3	AE003750
C 8	2131.6	88.2	75650	2	AC018145
9	962.2	39.8	4236	10	BC052433
10	960.6	39.7	2710	6	AX799084
11	955.8	39.5	2481	6	AX799082
12	955.8	39.5	2481	9	AY074880
13	733.4	30.3	2955	6	CQ573675
14	720.2	29.8	2608	5	BC046072
15	718.6	29.7	2567	5	BC063234
C 16	715.2	29.6	110000	8	CR382128_21
C 17	696	28.8	2953	6	CQ596819
18	636	26.3	2118	9	BT007100
19	636	26.3	2118	12	BT008132

20	636	26.3	2450	9	BC048348
21	636	26.3	2458	6	CQ834714
22	636	26.3	2472	6	CQ727867
23	636	26.3	2516	9	BC020965
24	636	26.3	2898	9	BSM809254
25	632.8	26.2	2293	9	HUMTMC
26	632.8	26.2	2472	6	AX799092
27	632.8	26.2	2472	9	HUMBSA
28	631.2	26.1	2736	10	BC037612
29	631.2	26.1	3094	6	AX799090
30	631.2	26.1	3094	10	MUSTRPR
31	628	26.0	2481	5	BC067313
32	628	26.0	3666	10	AX129027
33	623.2	25.8	2435	6	CQ834710
34	615.6	25.5	2760	6	CQ426624
35	615.6	25.5	2819	8	AK067609
36	614	25.4	2756	8	AK099138
37	607.4	25.1	1734	10	BC003206
38	582.2	24.1	54118	3	U13019
39	579.6	24.0	300029	8	AE016903
C 40	576.8	23.9	110000	8	CR382123_09
C 41	576.8	23.9	110000	8	CR382123_10
42	567.4	23.5	2555	8	BT012881
43	561.2	23.2	110000	8	CR382134_03
44	523.4	21.7	3633	8	SCYGL022W
45	521.8	21.6	2733	6	AX799086

ALIGNMENTS

RESULT 1	AF132552	Drosophila melanogaster	2417 bp	mRNA	linear	INV 12-FEB-2003
AF132552	Drosophila melanogaster	GM01838	full insert	cdNA.		
LOCUS	AF132552					
DEFINITION	AF132552					
ACCESSION	AF132552.1	GI:4689327				
VERSION	FLI CDNA.					
KEYWORDS	Drosophila melanogaster (fruit fly)					
SOURCE	Drosophila melanogaster					
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
REFERENCE	1 (bases 1 to 2417)					
AUTHORS	Tsang G., Brokstein, P., Frise, E., Harvey, D., Evans-Holm, M., Lewis, S.E., Suh, C. and Rubin, G.M.					
TITLE	Direct Submission					
JOURNAL	Submitted (02-MAR-1999) Berkeley Drosophila Genome Project, University of California Berkeley, 539 Life Sciences Addition #3200, Berkeley, CA 94720, USA					
REFERENCE	2 (bases 1 to 2417)					
AUTHORS	Stapleton, M., Brokstein, P., Hong, L., Agbavani, A., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Drensek, D., Farfan, D., Frise, E., George, R., Gonzalez, M., Guatin, H., Kronmiller, B., Li, P., Liao, G., Miranda, A., Mungall, C.J., Nuncio, J., Pacleb, J., Parggas, V., Park, S., Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M. and Celniker, S.					
TITLE	Direct Submission					
JOURNAL	Submitted (12-FEB-2003) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA					
COMMENT	Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory Berkeley, CA 94720					

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal

priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES

source

LOCATION
7 3417

I. :241/
/organism-"Dragonhila melanogaster"

mailto:info@openoffice.org

```

/mol_type="mRNA"
/ab_xref="+accn.7337"

```

Index

✓1. .2417

```
/gene="OstStt3"
```

$$\frac{d\mathbf{b}}{dt} = \mathbf{A}(\mathbf{b})\mathbf{b}$$

<1. ,2395

/gene="OstStt3"

```
/note="Longest ORF"
```

/codon_start=2

/product="GM01838p"

```

/feature key="protein id="AAD27851.2

```

```

/db xref="GI:28317200"

```

/db_xref="FUYBASE:FBgn0011336"
 /translation="LSEBVCVAFQGNFYRTWGNWDMRTPKMLNSKAVGSSYLTF
 AILLATLAWAGFSSRLFAVRESIEIHFFDFPNFYKAVYVQWGNKFNWVDEARAY
 PLGRIVGVGVPLGLMTSGGTHLLHLVNLPHVHIDICVLAPIBSGLTISYLLTK
 ELWSAGAGLPAASFTALVPGISRSVAGSDNEGIALPQFYFFVWRSYTGSGVFW
 SAAALSYVSVAGGVGVFINLIPUHFVLLMIGYSPRLTYSYTLFGLLFSN
 QYFVGFQFQRTISEMAALGVFVLMALVATEHLUQSVNRESERKFLVIGVLLGVGV
 FVAVVYTLGVVAPWSGRFSLMDTGAKIHPIIASVSEHQOTYFWFSGFDLHLV
 CAPPFGVVCYKIQINDESVFVILVAISAVFAGVYMWRLMVLTPVVCMLAGVAFSGLL
 FLDQBSKRMGTALSAATEVDEANESIEKTYLDKAGLTKHDKHQDQQTGVSSN
 LKSLVLBSMLMGTAHVCTVTSVSNYSVSLVLAHPSQDSRNLIDDFRAYVYL
 SQNTADRAWSWVDGYQIAGMARNTLVDNNWNHSHLAKVAMSKSTSEKYSIM
 TSLVDVLYIVFGGVYSGDIDNKFLMWVIAABEGRHQIKESDYFTDGRGVDAE
 GAPALLCNALMYKLSYVRGELKLDYSGSGVDYRTNVAIGNKFDLTLYEATVTEHW
 GAPALLCNALMYKLSYVRGELKLDYSGSGVDYRTNVAIGNKFDLTLYEATVTEHW

ORIGIN

Query Match	100.0%	Score	2417;	DB 3;	Length	2417;
Best Local Similarity	100.0%	Pred. No. 0;				
Matches 2417;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	TCTAAGCGAAGAAATGTGTCGTTGCAATTCAGATCGGTATATAATTTTCGAGTTACTGGCGT	60			
Db	1	TCTAAGCGAAGAAATGTGTCGTTGCAATTCAGATCGGTATATAATTTTCGAGTTACTGGCGT	60			
QY	61	GAATTGGACATGAATCGGACGCGGGAAGATCTGAACAGCAAGAGTGGCTGGCTACAGCAG	120			
Db	61	GAATTGGACATGAATCGGACGCGGGAAGATCTGAACAGCAAGAGTGGCTGGCTACAGCAG	120			
QY	121	CCTAATCACCTTCCGCATCTCTGTGAATCGCTGGCTGGCGCGAATTTTCTCTCGGCTCTT	180			
Db	121	CCTAATCACCTTCCGCATCTCTGTGAATCGCTGGCTGGCGCGAATTTTCTCTCGGCTCTT	180			
QY	181	GGCGGTCAATCGTTTCAGATCGAATATCCATGATGTTGATCCGTGGTTCAACTACCGGGC	240			
Db	181	GGCGGTCAATCGTTTCAGATCGAATATCCATGATGTTGATCCGTGGTTCAACTACCGGGC	240			
QY	241	CACCGCTACATGTCGCAGAAATGGTTGGTACAACTTCCCTCAACTGGTTCGACGAGCGGC	300			
Db	241	CACCGCTACATGTCGCAGAAATGGTTGGTACAACTTCCCTCAACTGGTTCGACGAGCGGC	300			
QY	301	ATGGTATCCGCTCGGCGAGGATGTGGGCGGTACCGTCTATCCCGGCTGTAGTTACGTC	360			
Db	301	ATGGTATCCGCTCGGCGAGGATGTGGGCGGTACCGTCTATCCCGGCTGTAGTTACGTC	360			
QY	361	CGGCGGAATCAATTGGCTGCTGCAGGTACTCAACATACACCGGTCATATTGGTGAATCTG	420			
Db	361	CGGCGGAATCAATTGGCTGCTGCAGGTACTCAACATACACCGGTCATATTGGTGAATCTG	420			
QY	421	CGTGTTCTCGGCGCGGATCTTTCAGTGGCTCACTCCATCTCCACTCACTCGTGAACAA	480			


```

QY 1561 GTTGATGATGTTGGTGTCTCCACTGCACTGGGTGACAGCAATGCCCTACTCCAGTCCCTC 1620
Db 1561 GTTGATGATGTTGGTGTCTCCACTGCACTGGGTGACAGCAATGCCCTACTCCAGTCCCTC 1620
QY 1621 CATTGTCTTGCTGTTTCCCAACAGTCAAGATGGATCCCGCAACATTTTACAGCATTTTCAG 1680
Db 1621 CATTGTCTTGCTGTTTCCCAACAGTCAAGATGGATCCCGCAACATTTTACAGCATTTTCAG 1680
QY 1681 AGAGGCTTACTCTGGCTTTCCAGAACTGCGCGATGATCTCGCGTTATGCTTGGTG 1740
Db 1681 AGAGGCTTACTCTGGCTTTCCAGAACTGCGCGATGATCTCGCGTTATGCTTGGTG 1740
QY 1741 CGATTACGGATACACAGATAGCGGAAATGCGAAACAGAACGACGCTAGTGGATATATATAC 1800
Db 1741 CGATTACGGATACACAGATAGCGGAAATGCGAAACAGAACGACGCTAGTGGATATATATAC 1800
QY 1801 GTGGAAACATAGTCACTAGCGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1860
Db 1801 GTGGAAACATAGTCACTAGCGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1860
QY 1861 CTACGAAATATGACATCTCTGACCTGGACTAGCTTGTGTTGTTGATCTTTGGCGGTGTGAT 1920
Db 1861 CTACGAAATATGACATCTCTGACCTGGACTAGCTTGTGTTGTTGATCTTTGGCGGTGTGAT 1920
QY 1921 CGGCTATTCTGCGATGATATCAACAGTCTCTGATGATGATGATGATGATGATGATGATGAT 1980
Db 1921 CGGCTATTCTGCGATGATATCAACAGTCTCTGATGATGATGATGATGATGATGATGATGAT 1980
QY 1981 GCATCCCAAGGACATTAAGGAAAGGATTAATTTACCGACCGCGGTGAAATTCAGGGTAGA 2040
Db 1981 GCATCCCAAGGACATTAAGGAAAGGATTAATTTACCGACCGCGGTGAAATTCAGGGTAGA 2040
QY 2041 TGCCGAAGGTGCTCGCGGCTGCTCAACTGCTTATGTAACAAATTAAGCTACTACAGATT 2100
Db 2041 TGCCGAAGGTGCTCGCGGCTGCTCAACTGCTTATGTAACAAATTAAGCTACTACAGATT 2100
QY 2101 CGCGGAATTAAGTTGGACTACAGAGTCCATCTGGATATGATCGCACACGTAACGCGCT 2160
Db 2101 CGCGGAATTAAGTTGGACTACAGAGTCCATCTGGATATGATCGCACACGTAACGCGCT 2160
QY 2161 CATTGGGAATAAGACTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 2220
Db 2161 CATTGGGAATAAGACTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 2220
QY 2221 GCTTGTTCGATCTATAGGTGAAGAGCCGATGATTAATAGACCATCACTGGAAGAC 2280
Db 2221 GCTTGTTCGATCTATAGGTGAAGAGCCGATGATTAATAGACCATCACTGGAAGAC 2280
QY 2281 CAAGGAGAGACGATTCCTCCAGCAAACTTCAATTCGAGAAAGAACTCTTAAGCGTGGCAA 2340
Db 2281 CAAGGAGAGACGATTCCTCCAGCAAACTTCAATTCGAGAAAGAACTCTTAAGCGTGGCAA 2340
QY 2341 GGGCTACATACGAAACCGACCGGTTGTTGTTAAGGAAACGAAACCTTGAAATAAACCCA 2400
Db 2341 GGGCTACATACGAAACCGACCGGTTGTTGTTAAGGAAACGAAACCTTGAAATAAACCCA 2400
QY 2401 AAAAAAATAAAAAA 2417
Db 2401 AAAAAAATAAAAAA 2417

```

```

RESULT 2
AX799088 2417 bp DNA linear PAT 08-OCT-2003
LOCUS Sequence 7 from Patent WO03054008.
DEFINITION AX799088
ACCESSION AX799088
VERSION AX799088.1 GI:37605061
KEYWORDS Drosophila melanogaster (fruit fly)
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```

```

Ephydroidea; Drosophilidae; Drosophila.
1
Perreault, C. and McBride, K.
Mammalian SIMP protein, gene sequence and uses thereof in cancer
therapy
Patent: WO 03054008-A 7 03-JUL-2003;
Compatigene Inc. (CA)
Location/Qualifiers
1. 2417
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/note="STT3, Acc# AF132552"

ORIGIN
Query Match 100.0%; Score 2417; DB 6; Length 2417;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAAGCGAAGAATGTGTGCTTGCATTTTCAGATCGGTTAATAATTTTCGAGTTACTGGCTG 60
Db 1 TCTAAGCGAAGAATGTGTGCTTGCATTTTCAGATCGGTTAATAATTTTCGAGTTACTGGCTG 60
QY 61 GAATTGGGACATGAATCGGACGCGAGATGCTGACGCAAGTGGCTGCTACAGCAG 120
Db 61 GAATTGGGACATGAATCGGACGCGAGATGCTGACGCAAGTGGCTGCTGCTACAGCAG 120
QY 121 CCTAATCACTTCGCCATCTCTGCTAATTCGGCTGGCTGGCGGATTTTCTCTCGCCTCTT 180
Db 121 CCTAATCACTTCGCCATCTCTGCTAATTCGGCTGGCTGGCGGATTTTCTCTCGCCTCTT 180
QY 181 CGCGGTATCTCGTTTCGATGATTCATGATGATGATGATGATGATGATGATGATGATGATG 240
Db 181 CGCGGTATCTCGTTTCGATGATTCATGATGATGATGATGATGATGATGATGATGATGATG 240
QY 241 CACGCGCTACATGTGTGCGAATGCTTGTGTAACATCTTCTCAACTGTTTCGACGAGCGCG 300
Db 241 CACGCGCTACATGTGTGCGAATGCTTGTGTAACATCTTCTCAACTGTTTCGACGAGCGCG 300
QY 301 ATGGTATCTCGCTCGGAGGATTTGGCGGTACCGTCTATCCGGCTGATGATGATGATGATG 360
Db 301 ATGGTATCTCGCTCGGAGGATTTGGCGGTACCGTCTATCCGGCTGATGATGATGATGATG 360
QY 361 CGGCGGAATCAATTTGGTGTGCTGACATCAACATACCGGTCCATATTCGTGACATCTG 420
Db 361 CGGCGGAATCAATTTGGTGTGCTGACATCAACATACCGGTCCATATTCGTGACATCTG 420
QY 421 CGTGTTCCTGCGCGGATCTTCAGTGGCTGACCTCACTCCACCTACTGCTGACCA 480
Db 421 CGTGTTCCTGCGCGGATCTTCAGTGGCTGACCTCACTCCACCTACTGCTGACCA 480
QY 481 GGAGCTGTGTGCTCGCGCGGCGCGGCTTCTTCGCGCGGAGCTTCATCGCCATCGTGCCTG 540
Db 481 GGAGCTGTGTGCTCGCGCGGCGCGGCTTCTTCGCGCGGAGCTTCATCGCCATCGTGCCTG 540
QY 541 CTACATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Db 541 CTACATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
QY 601 GCAGTTCACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Db 601 GCAGTTCACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
QY 661 CGCAGCGCTTGTCTTACTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 661 CGCAGCGCTTGTCTTACTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 721 CAACCTGATACCTCTGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 721 CAACCTGATACCTCTGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 781 GCTGACGAGCTACAGACCTTCTACATCTCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840

```


Db	70	GAATTGGGACATGAATCGACGCCGAGATGCTGAACAGCAAGGTGGCTGGCTACAGCAG	122
QY	121	CCTAATCACCTTCGCCCATCCTGCTAATCGCCTGGCTGGCGGATTTCTCTCGCCTCTT	180
Db	130	CCTAATCACCTTCGCCCATCCTGCTAATCGCCTGGCTGGCGGATTTCTCTCGCCTCTT	189
QY	181	CSGCGTCATCCGTTTCGAGTCGATTATCATGAGTTTGATCGTGTTGATCGTGTTCAACTACCGGCG	240
Db	190	CGCGGTCACTCCGTTTCGAGTCGATTATCATGAGTTTGATCGTGTTGATCGTGTTCAACTACCGGCG	249
QY	241	CACCGCCTACATGTTGTCAGAAATGGTTGGTACAACTTCCTCACTGGTTCGACGAGCGGCG	300
Db	250	CACCGCCTACATGTTGTCAGAAATGGTTGGTACAACTTCCTCACTGGTTCGACGAGCGGCG	309
QY	301	ATGGTATCCGTCGCGCAGAGATTGTGGGGGTACCGTCTATCCGGCCTGATGATTACGTC	360
Db	310	ATGGTATCCGTCGCGCAGAGATTGTGGGGGTACCGTCTATCCGGCCTGATGATTACGTC	369
QY	361	CGCGCGAATCCATTGGCTGCTGCAGTACTCAACTACGGGTCCATATCGTGACATCTG	420
Db	370	CGCGCGAATCCATTGGCTGCTGCAGTACTCAACTACGGGTCCATATCGTGACATCTG	429
QY	421	CGTGTCTTCGCGCGGATCTTCAGTGGCTGACCTCCATCTCCCACTACCTGCTGACCAA	480
Db	430	CGTGTCTTCGCGCGGATCTTCAGTGGCTGACCTCCATCTCCCACTACCTGCTGACCAA	489
QY	481	GGAGCTGTGGTCCGGGGCGCGGCTCTTCGCGCGCAGCTTCATCGGCATCTGCGCTGG	540
Db	490	GGAGCTGTGGTCCGGGGCGCGGCTCTTCGCGCGCAGCTTCATCGGCATCTGCGCTGG	549
QY	541	CTACATCAGTAGTGGTGGCTGGATCGTACGATAACGAGGGCATTTGCCATATTCCGCCCT	600
Db	550	CTACATCAGTAGTGGTGGCTGGATCGTACGATAACGAGGGCATTTGCCATATTCCGCCCT	609
QY	601	GCAGTTCACCTACTTCTGTGGGTGGCTCAGTGAAGACTGGATCCGTTGTTCTGGTGGCG	660
Db	610	GCAGTTCACCTACTTCTGTGGGTGGCTCAGTGAAGACTGGATCCGTTGTTCTGGTGGCG	669
QY	661	CGCAGCGCTTTGTCTACTTCTACATGGTTCGGCCTGGGGTGGCTACGTTTTCATCAT	720
Db	670	CGCAGCGCTTTGTCTACTTCTACATGGTTCGGCCTGGGGTGGCTACGTTTTCATCAT	729
QY	721	CAACCTGATACCCCTGCACGCTTCGTACTGCTCATTTAGGGGAGTACTCGCGCGCT	780
Db	730	CAACCTGATACCCCTGCACGCTTCGTACTGCTCATTTAGGGGAGTACTCGCGCGCT	789
QY	781	GCTCACCAGCTACAGCACCTTCTACATCCCTGGGACTGCTGTTCCTCCATGAGATCCCTT	840
Db	790	GCTCACCAGCTACAGCACCTTCTACATCCCTGGGACTGCTGTTCCTCCATGAGATCCCTT	849
QY	841	CGTGGGATTCCAAACCGATAACGACAGTGAACATAGGCTGCGTGGGAGTTTGTGCT	900
Db	850	CGTGGGATTCCAAACCGATAACGACAGTGAACATAGGCTGCGTGGGAGTTTGTGCT	909
QY	901	CCTTATGGCGGTGGCCACCTTTCGCCATTTGACGTCGCTGTCGCGCAACGAGTTCG	960
Db	910	CCTTATGGCGGTGGCCACCTTTCGCCATTTGACGTCGCTGTCGCGCAACGAGTTCG	969
QY	961	GAAGCTGTTTCATCGTCGGCGGATTTGTTGGGCGTTGGCGTCTTTTGTGGCGCTCGTGGT	1020
Db	970	GAAGCTGTTTCATCGTCGGCGGATTTGTTGGGCGTTGGCGTCTTTTGTGGCGCTCGTGGT	1029
QY	1021	GCTCACCATCTGGGCGTTGGCCCCGTGGAGTGGAGCGTTCTACTCGTGTGGGATAC	1080
Db	1030	GCTCACCATCTGGGCGTTGGCCCCGTGGAGTGGAGCGTTCTACTCGTGTGGGATAC	1089
QY	1081	TGGCTACGCCAAGATCCACATTCCTCATTTGATCCGTCGAGGATCAGGCCACCAC	1140
Db	1090	TGGCTACGCCAAGATCCACATTCCTCATTTGATCCGTCGAGGATCAGGCCACCAC	1149
QY	1141	TTGGTCTCTGTTCTTTTGTGATCTGACATCTGTTGGTGGCGCTTCCAGTGGGAGTGTG	1200
Db	1150	TTGGTCTCTGTTCTTTTGTGATCTGACATCTGTTGGTGGCGCTTCCAGTGGGAGTGTG	1209

QY 2281 CAAGGAGAGACGATTCTCCAGCAAACTTCATTTTCGAGAAAGAACTTAAGCGTCGCAA 2340
 Db 2290 CAAGGAGAGACGATTCTCCAGCAAACTTCATTTTCGAGAAAGAACTTAAGCGTCGCAA 2349
 QY 2341 GGGCTACATACGAAACCGACCGGTTGTTTAAAGGAAAGAAAGAACTTGAATAAAGCCCA 2400
 Db 2350 GGGCTACATACGAAACCGACCGGTTGTTTAAAGGAAAGAAAGAACTTGAATAAAGCCCA 2409
 QY 2401 AAAAAAATAA 2412
 Db 2410 AAAAAAATAA 2421
 RESULT 4
 CQ589352/c
 LOCUS
 DEFINITION Sequence 17110 from Patent WO0171042.
 ACCESSION CQ589352
 VERSION CQ589352.1 GI:41648214
 KEYWORDS
 SOURCE Drosophila sp.
 ORGANISM Drosophila sp.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1
 Venter, J.C.; Adams, M., Li, P.W. and Myers, E.W.
 Detection kits, such as nucleic acid arrays, for detecting the
 expression of 10,000 or more Drosophila genes and uses thereof
 Patent: WO 0171042-A 17110 27-SEP-2001;
 PE Corporation (NY) (US)
 FEATURES
 Location/Qualifiers
 1..4922
 /organism="Drosophila sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:7242"
 ORIGIN
 Query Match 88.8%; Score 2147.4; DB 6; Length 4922;
 Best Local Similarity 91.5%; Pred. No. 0;
 Matches 2411; Conservative 0; Mismatches 1; Indels 223; Gaps 4;
 QY 1 TCTAAGCGAAGATGTCGTTGCATTTCAGATTCGGTATAATTTTCGAGTTACTGGCTG 60
 Db 3913 TCTAAGCGAAGATGTCGTTGCATTTCAGATTCGGTATAATTTTCGAGTTACTGGCTG 3854
 QY 61 GAATTGGGACATGAATCGGACGCCGGAAGATGCTGAACAGCAAGTGGCTGCTACAGCAG 120
 Db 3853 GAATTGGGACATGAATCGGACGCCGGAAGATGCTGAACAGCAAGTGGCTGCTACAGCAG 3794
 QY 121 CCTAATCACCTTCGACATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTT 180
 Db 3793 CCTAATCACCTTCGACATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTT 3734
 QY 181 CGCGCTCATCCGTTTCGAGTCGATTAATCCATGAGTTTGAATCCGTTGATCACTACCGGC 240
 Db 3733 CGCGCTCATCCGTTTCGAGTCGATTAATCCATGAGTTTGAATCCGTTGATCACTACCGGC 3674
 QY 241 CACGGCTACATGGTCGAGATGTTGTTGACACTTCCCTCAACTGGTTCGACGAGCGGC 300
 Db 3673 CACGGCTACATGGTTCGAGATGTTGTTGACACTTCCCTCAACTGGTTCGACGAGCGGC 3614
 QY 301 ATGTATCCGCTCGGACGAGTTGCGGCGTACCGCTATCCCGGCTCATGATACGTC 360
 Db 3613 ATGTATCCGCTCGGACGAGTTGCGGCGTACCGCTATCCCGGCTCATGATACGTC 3554
 QY 361 CCGCGGAATCATGGTTCGCTGACGCTACATACCGCTCATATTCGTCGATCTG 420
 Db 3553 CCGCGGAATCATGGTTCGCTGACGCTACATACCGCTCATATTCGTCGATCTG 3494
 QY 421 CGTGTTCCTGGCGCGGATCTTCAGTGGCCTGACCTCCATCTCCACCTACTGTCGACCAA 480
 Db 3493 CGTGTTCCTGGCGCGGATCTTCAGTGGCCTGACCTCCATCTCCACCTACTGTCGACCAA 3434

QY 481 GGAGCTGTGGTCCGCGGCGCGGCTCTTCCGCGCAGCTTTCATCGCCATCGTGCCTGG 540
 Db 3433 GGAGCTGTGGTCCGCGGCGCGGCTCTTCCGCGCAGCTTTCATCGCCATCGTGCCTGG 3374
 QY 541 CTACATCAGTAGTTCGCTGGCTGATCGTACGATAACGAGGCGATTGCGCATATTTCGCGCT 600
 Db 3373 CTACATCAGTAGTTCGCTGGCTGATCGTACGATAACGAGGCGATTGCGCATATTTCGCGCT 3314
 QY 601 GCAGTTCACTACTTCTCTGTGGTTCGCTCAGTGAAGACTGATCCGTTGTTCTGGTGGC 660
 Db 3313 GCAGTTCACTACTTCTCTGTGGTTCGCTCAGTGAAGACTGATCCGTTGTTCTGGTGGC 3254
 QY 661 CGAGCGCTTTTCTCTACTTCTACATGGTTCGCTCCGCTGGGCTGCTACGTTGTTTCATCAT 720
 Db 3253 CGAGCGCTTTTCTCTACTTCTACATGGTTCGCTCCGCTGGGCTGCTACGTTGTTTCATCAT 3194
 QY 721 CAACCTGATACCCCTGCACGCTCTTCTGATCTGCTATTATGGCAGGTACTCGCGCGTCT 780
 Db 3193 CAACCTGATACCCCTGCACGCTCTTCTGATCTGCTATTATGGCAGGTACTCGCGCGTCT 3134
 QY 781 GCTGACCACTACAGCACTTCTACATCTGGGACTGCTTCTCCATCGAGATCCCTCT 840
 Db 3133 GCTGACCACTACAGCACTTCTACATCTGGGACTGCTTCTCCATCGAGATCCCTCT 3074
 QY 841 CGTGGGATTCCAAACCGATACGACAGTGAACACATGCTGCGCTGGGAGTGTTCGCT 900
 Db 3073 CGTGGGATTCCAAACCGATACGACAGTGAACACATGCTGCGCTGGGAGTGTTCGCT 3014
 QY 901 CTTATGGCGGTGGCCACTTTCGCGCATTTGAGTCCGCTGCTGCGGCAAGAGTTCGG 960
 Db 3013 CTTATGGCGGTGGCCACTTTCGCGCATTTGAGTCCGCTGCTGCGGCAAGAGTTCGG 2954
 QY 961 GAAGCTGTTTCATGCTCGGCGGATTTGCTGCTGGCGTTCGCTGCTTTCGCGCTGCTGCT 1020
 Db 2953 GAAGCTGTTTCATGCTCGGCGGATTTGCTGCTGGCGTTCGCTGCTTTCGCGCTGCTGCT 2894
 QY 1021 GCTCACCATGCTGGCGGTTTGGCCCGGTGGAGTGAAGCTTCTACTCTGCTGGGATAC 1080
 Db 2893 GCTCACCATGCTGGCGGTTTGGCCCGGTGGAGTGAAGCTTCTACTCTGCTGGGATAC 2834
 QY 1081 TGCTACGCCAAGATCCATTCCTCCATTCATCCATCCGCTGCGAGCATACGCCACAC 1140
 Db 2833 TGCTACGCCAAGATCCATTCCTCCATTCATCCATCCGCTGCGAGCATACGCCACAC 2774
 QY 1141 TTGGTTCCTGCTTCTTTTTCATCTGCACATCTCTGCTGCTGCTTCCAGTGGGAGTGTG 1200
 Db 2773 TTGGTTCCTGCTTCTTTTCATCTGCACATCTCTGCTGCTGCTTCCAGTGGGAGTGTG 2714
 QY 1201 GTACTGATCAAGCAGATCAACGACGAGCGCTTTTCGCTGCTGCTGCTGCTGCTGCTG 1260
 Db 2713 GTACTGATCAAGCAGATCAACGACGAGCGCTTTTCGCTGCTGCTGCTGCTGCTGCTG 2654
 QY 1261 GGTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
 Db 2653 GGTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2594
 QY 1321 GCTGGCGGAGTGGCTTTTCGCGACTGTTGGATGTTGCTGCTGCTGCTGCTGCTGCTG 1380
 Db 2593 GCTGGCGGAGTGGCTTTTCGCGACTGTTGGATGTTGCTGCTGCTGCTGCTGCTGCTG 2534
 QY 1381 GCGAATCGGACAGCCATAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 1440
 Db 2533 GCGAATCGGACAGCCATAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 2474
 QY 1441 GAAGAAGACGCTGTACACA ----- 1460
 Db 2473 GAAGAAGACGCTGTACACAAGGTGAGTTCTTACTAACAACATCCATCCATGGGATTTGATTT 2414
 QY 1461 -----AGGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG 1501
 Db 2413 TAAATACATCGCATTTGCGGCTGGCAAGCTGAGCATCTGTAAGCATGATGCCAG 2354

```
QY 1502 CAGGATACCTGGCGTCCAGCTCAACCTGAAGAGTATGTTATTTTGGCGCTTCATATGCTG 1561
D 2353 CAGGATACCTGGCGTCCAGCTCAACCTGAAGAGTATGTTATTTTGGCGCTTCATATGCTG 2294
QY 1562 TTGATGATGTTGCGTGTCCACTGACGCTGGGTGACCAAGCAATGCTACTCCAGTCCCTCC 1621
D 2293 TTGATGATGTTGCGTGTCCACTGACGCTGGGTGACCAAGCAATGCTACTCCAGTCCCTCC 2234
QY 1622 ATTGCTTGGCTTTCCCAACAGTCAAGATG----- 1652
D 2233 ATTGCTTGGCTTTCCCAACAGTCAAGATGGGTAAAGCACCAGATCATGCTGTTCT 2174
QY 1653 -----GATCCCGCAACATTTTAGACGATTTTCAGAGAG 1684
D 2173 TTAAACCTATCTAACCTTTTCTCCATAGATCCCGCAACATTTTAGACGATTTTCAGAGAG 2114
QY 1685 GCTTACTACTGCTTTCCGAGAACATCTGCGATGATGCTGCGTATATGTTCTTGGTGGAT 1744
D 2113 GCTTACTACTGCTTTCCGAGAACATCTGCGATGATGCTGCGTATATGTTCTTGGTGGAT 2054
QY 1745 TACGATACCATAGCGGATGCGAAACAGACGAGCTAGTGGGATATATATACGTGG 1804
D 2053 TACGATACCATAGCGGATGCGAAACAGACGAGCTAGTGGGATATATATACGTGG 1994
QY 1805 AACATAGTACATAGCTGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTAC 1864
D 1993 AACATAGTACATAGCTGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTAC 1934
QY 1865 GAAATATGATCTCTTACGCTGACCTACGTTTGGTGTATCTTTGCGGTGATCGGC 1924
D 1933 GAAATATGATCTCTTACGCTGACCTACGTTTGGTGTATCTTTGCGGTGATCGGC 1874
QY 1925 TATTCGCGATGATATCAACAAAGTCTCTGTGGATGCTGCGAATGCTGAGGAGAGCAT 1984
D 1873 TATTCGCGATGATATCAACAAAGTCTCTGTGGATGCTGCGAATGCTGAGGAGAGCAT 1814
QY 1985 CCAAGGACATTAAGGAAGCGATTACTTTACCAACCGCGTGAATTCAGGCTAGATGCC 2044
D 1813 CCAAGGACATTAAGGAAGCGATTACTTTACCAACCGCGTGAATTCAGGCTAGATGCC 1754
QY 2045 GAAGTGTCTCGGCGCTCTCAACTGCTGTATGATACAAATTAAGCTACTACAGATTCGGG 2104
D 1753 GAAGTGTCTCGGCGCTCTCAACTGCTGTATGATACAAATTAAGCTACTACAGATTCGGG 1694
QY 2105 GAATGAATGTTGGACTA----- 2121
D 1693 GAATGAATGTTGGACTACAGTAAGCGGAACATTTCTCAGGTAGCGATGCACACTAA 1634
QY 2122 -----CAGAGTCCATCTGGATATGATCGCACACGTAAACCGCGTCAATTTGGGAATA 2171
D 1633 CTGCTGTTTTCAGAGGTCCATCTGGATATGATCGCACACGTAAACCGCGTCAATTTGGGAATA 1574
QY 2172 AGGACTTCGATCTGACCTACCTGGAGGAGGCTACACCACAGACACTGCGTGTTCGCA 2231
D 1573 AGGACTTCGATCTGACCTACCTGGAGGAGGCTACACCACAGACACTGCGTGTTCGCA 1514
QY 2232 TCTATAGGTGAAGACCGCATGATGTTCAATAGACCATCACTGAAGACCAAGAGAGAA 2291
D 1513 TCTATAGGTGAAGACCGCATGATGTTCAATAGACCATCACTGAAGACCAAGAGAGAA 1454
QY 2292 CGATTCCTCCAGCAACCTTCATTTTCGAGAA----- 2321
D 1453 CGATTCCTCCAGCAACCTTCATTTTCGAGAAAGGTAGTGTACCTCGAGCTCCCAATGAAC 1394
QY 2322 -----AGAACTTAAGCGTCCGCAAGGGCTACATACGAAACC 2357
D 1393 TTCATTAACAAATCTGATTTCTTTGAGAACTCTAAGCGTCCGCAAGGGCTACATACGAAACC 1334
QY 2358 GACCGGTGTTGTTAAGGAGAAACGACCTTGAATATAACCCAAAAAATAA 2412
D 1333 GACCGGTGTTGTTAAGGAGAAACGACCTTGAATATAACCCAAAAAATAA 1279
```

RESULT 5
AC007853/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches 2411

Conservative

Mismatches

Indels

Gaps

QY

AC007853 162921 bp DNA linear INV 27-FEB-2001

Drosophila melanogaster, chromosome 3R, region 96B-96C, BAC clone

BACR03L02, complete sequence.

AC007853 4 GI:13129410

HTG.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 162921)

Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,

Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,

Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,

Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,

Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Drenek,D., Farfan,D.,

Fierera,S., Frise,E., Galle,R.F., Gaig,N.S., George,R.A.,

Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,

Ibegwan,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,

McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nuncio,J.,

Pacib,J., Paragas,V., Park,S., Patel,S., Pfeiffer,S.,

Phanphanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,

Shapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,

Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

Sequencing of Drosophila chromosome 3R, region 96B-96C

Unpublished

2 (bases 1 to 162921)

Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,

Butenhoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,

Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,

Hoskins,R., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,

Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacib,J.M., Park,S.,

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,

Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and

Rubin,G.M.

Direct Submission

Submitted (17-JUN-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 27, 2001 this sequence version replaced gi:5670481.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome

shotgun and from subclones of this BAC and its neighboring clones.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send emailto bdgp@fruitfly.berkeley.edu.

Location/Qualifiers

1..162921

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/strain="y; cn bw sp"

/db_xref="taxon:7227"

/chromosome="3R"

/map="96B-96C"

/clone="BACR03L02 (D766)"

/clone_lib="RPCI-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

pBACe3.6)"

ORIGIN

Query Match 88.8% Score 2147.4; DB 3; Length 162921;

Best Local Similarity 91.58; Pred No. 0;

Matches 2411; Conservative 0; Mismatches 1; Indels 223; Gaps 4;

QY 1 TCTAAGCGAAGATGCTCGTTCATTCAGATCGGTATATATTTTCGAGTACTGGCTG 60

Db 108551 TCTAAGCGAAGATGTGCTGCTCATTTTCAGATCGGTTATAATTTTCGAGTTACTGGCTG 108492
Qy 61 GAATGGGACATGAATCGGACGCGGAGATGCTGAACAGCAGGTGGCTGGCTACAGCAG 120
Db 108491 GAATGGGACATGAATCGGACGCGGAGATGCTGAACAGCAGGTGGCTGGCTACAGCAG 108432
Qy 121 CCTAATCACCTTCGCCATCCTGCTAATCGCTGGCTGGCGGAGATTTCTCTCGCTCTT 180
Db 108431 CCTAATCACCTTCGCCATCCTGCTAATCGCTGGCTGGCGGAGATTTCTCTCGCTCTT 108372
Qy 181 GCGCTCATCGTTTCGAGTCGATTAATCCATGAGTTTGAATCGTGGTTCACTACCGGCG 240
Db 108371 GCGCTCATCGTTTCGAGTCGATTAATCCATGAGTTTGAATCGTGGTTCACTACCGGCG 108312
Qy 241 CACCGCTACATGCTGACGAATGTTGGTACAACTTCTCAACTGGTTTCGACGAGCGCG 300
Db 108311 CACCGCTACATGCTGACGAATGTTGGTACAACTTCTCAACTGGTTTCGACGAGCGCG 108252
Qy 301 ATGCTATCGCTCGGACGAGTTGTGGCGGTACCGTCTATCCGCGCTGATGATTAAGTC 360
Db 108251 ATGCTATCGCTCGGACGAGTTGTGGCGGTACCGTCTATCCGCGCTGATGATTAAGTC 108192
Qy 361 CGCGGGAATCCATGCTGCTGACGCTACTCAACATACCGGTCCCATATTCGTGACATCTG 420
Db 108191 CGCGGGAATCCATGCTGCTGACGCTACTCAACATACCGGTCCCATATTCGTGACATCTG 108132
Qy 421 CGTGTTCCTGGCGCGATCTTCACTGGCTGACCTCCATCTCCACTACCTGCTGACCAA 480
Db 108131 CGTGTTCCTGGCGCGATCTTCACTGGCTGACCTCCATCTCCACTACCTGCTGACCAA 108072
Qy 481 GGAGCTGTGCTCGCGCGCGCGCTCTTTCGCGCGAGCTTCATCGCCTGCTGCTGG 540
Db 108071 GGAGCTGTGCTCGCGCGCGCGCTCTTTCGCGCGAGCTTCATCGCCTGCTGCTGG 108012
Qy 541 CTACATCAGTAGTTCGCTGCTGATCGTACGATACGAGGCGATTCGCTATTCGCT 600
Db 108011 CTACATCAGTAGTTCGCTGCTGATCGTACGATACGAGGCGATTCGCTATTCGCT 107952
Qy 601 GCAGTTCACCTACTTCTGCTGGTGGCTCAGTGAAGACTGGATCCGTTGTTCTGGTCGCG 660
Db 107951 GCAGTTCACCTACTTCTGCTGGTGGCTCAGTGAAGACTGGATCCGTTGTTCTGGTCGCG 107892
Qy 661 CGAGCGCTTCTCTACTTCTACATGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 720
Db 107891 CGAGCGCTTCTCTACTTCTACATGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 107832
Qy 721 CAACCTGATACCCCTGACGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 107831 CAACCTGATACCCCTGACGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 107772
Qy 781 GCTGACGAGTACGACCTTCTACATCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 107771 GCTGACGAGTACGACCTTCTACATCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 107712
Qy 841 CGTGGGATCCAAACCGATACGACCGAGTGAACACATGGCTGGCTGGAGTGTGCTGCT 900
Db 107711 CGTGGGATCCAAACCGATACGACCGAGTGAACACATGGCTGGCTGGAGTGTGCTGCT 107652
Qy 901 CCTTATGGCGTGGCGACCTTGGCGCATTTGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 107651 CCTTATGGCGTGGCGACCTTGGCGCATTTGAGTCCGCTGCTGCTGCTGCTGCTGCTGCT 107592
Qy 961 GAAGCTGTTTCATCGTGGCGGATTTGCTGGCGGCTGGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 107591 GAAGCTGTTTCATCGTGGCGGATTTGCTGGCGGCTGGCTGCTGCTGCTGCTGCTGCTGCT 107532
Qy 1021 GCTCACCATCTGGCGGTTGTGGCGCGGTGGAGTGGAGCTTCTACTGCTGCTGGATAC 1080
Db 107531 GCTCACCATCTGGCGGTTGTGGCGCGGTGGAGTGGAGCTTCTACTGCTGCTGGATAC 107472
Qy 1081 TGGCTACGCCAAGATCCACATTCCTCATTCATTCGCTGCTGGAGCATCAGCCACAC 1140
Db 107471 TGGCTACGCCAAGATCCACATTCCTCATTCATTCGCTGCTGGAGCATCAGCCACAC 107412

Qy 1141 TTGGTTCCTGCTTCTTCTTGTATCTGCACATCCTGCTGCGCCTTCCAGTGGAGTGTG 1200
Db 107411 TTGGTTCCTGCTTCTTCTTGTATCTGCACATCCTGCTGCGCCTTCCAGTGGAGTGTG 107352
Qy 1201 GTACTGCTCAAGCAGATCAACGACGAGCGGTTTCTGCTGCTGTGTACCCATCAGTGC 1260
Db 107351 GTACTGCTCAAGCAGATCAACGACGAGCGGTTTCTGCTGCTGTGTGTACCCATCAGTGC 107292
Qy 1261 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 107291 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 107232
Qy 1321 GCTGCGGAGTGGCTTTTTCGGGACTGTTGGATGTTGTTCTGCTCAAGAGATTCGCTAA 1380
Db 107231 GCTGCGGAGTGGCTTTTTCGGGACTGTTGGATGTTGTTCTGCTCAAGAGATTCGCTAA 107172
Qy 1381 GCGAATGGGCGACGATCAAGCAGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
Db 107171 GCGAATGGGCGACGATCAAGCAGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 107112
Qy 1441 GAAGAAGCGCTGTACGACA----- 1460
Db 107111 GAAGAAGCGCTGTACGACAAGGTGAGTTCTTACTAACACACATCCATGGGATTTGATTT 107052
Qy 1461 ----- 1501
Db 107051 TAAATATACATGCTATTCGAGCTGGGAGCTGAGCATTTGTAAGCATGTAAGCATGTAAGCAT 106992
Qy 1502 CAGGATATCGGCTCAGCTCCAACTGAAAGATTTGTTATTTTGGCCGTTCTTAATGCTG 1561
Db 106991 CAGGATATCGGCTCAGCTCCAACTGAAAGATTTGTTATTTTGGCCGTTCTTAATGCTG 106932
Qy 1562 TTGATGATGCTGCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1621
Db 106931 TTGATGATGCTGCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 106872
Qy 1622 ATTGCTTGGCTTTCACAAACAGTCAAGATG----- 1652
Db 106871 ATTGCTTGGCTTTCACAAACAGTCAAGATGTTGTTGTTGTTGTTGTTGTTGTTGTT 106812
Qy 1653 ----- 1684
Db 106811 TTAACCTATCTAACCTTTTCTCCACTAGATCCCGCAACATTTTAGACGATTTACAGAG 106752
Qy 1685 GCTTACTGCTGCTTTCGAAACACTGCGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1744
Db 106751 GCTTACTGCTGCTTTCGAAACACTGCGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 106692
Qy 1745 TACGATACCGATAGCGGATGCGAAACAGACAGCAGCTAGTGGATATAATACGTTG 1804
Db 106691 TACGATACCGATAGCGGATGCGAAACAGACAGCAGCTAGTGGATATAATACGTTG 106632
Qy 1805 AACATAGTCAATAGCCTGTTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTAC 1864
Db 106631 AACATAGTCAATAGCCTGTTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTAC 106572
Qy 1865 GAAATATGACATCTTTCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1924
Db 106571 GAAATATGACATCTTTCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 106512
Qy 1925 TATTTCGCGATGATPATAACAAAGTTCCTGTTGATGTTGCGAATTTGCTGAGGAGAGCAT 1984
Db 106511 TATTTCGCGATGATPATAACAAAGTTCCTGTTGATGTTGCGAATTTGCTGAGGAGAGCAT 106452
Qy 1985 CCAAGGACATTAAGGAGCGATTTACTTTTACCGACCGGCTGAAATTCAGGCTAGATGCC 2044
Db 106451 CCAAGGACATTAAGGAGCGATTTACTTTTACCGACCGGCTGAAATTCAGGCTAGATGCC 106392
Qy 2045 GAAGTGTCTCGGCGCTGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2104
Db 106391 GAAGTGTCTCGGCGCTGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 106332

431	QY	CGTGTTCCTGGCGCCGATCTTCAGTGGCGCTGACCTCCATCTCCACTACCTGCTGACAA	480
130650	DB	CGTGTTCCTGGCGCCGATCTTCAGTGGCGCTGACCTCCATCTCCACTACCTGCTGACAA	130591
481	QY	GGAGCTGTGGTCCGCGGGCGCGGCTCTTCGCGCGCAGCTTCATCGCCATCGTGCGCTGG	540
130590	DB	GGAGCTGTGGTCCGCGGGCGCGGCTCTTCGCGCGCAGCTTCATCGCCATCGTGCGCTGG	130531
541	QY	CTACATCAGTAGTTCGTGGCTGGATCGTAGATTAACGAGGCGATTCGCATATTCGCGCT	600
130530	DB	CTACATCAGTAGTTCGTGGCTGGATCGTAGATTAACGAGGCGATTCGCATATTCGCGCT	130471
601	QY	GCAGTTCACGCTACTTCCTGTGGGTGGGCTCAGTAGAGACTGGATCCGTGTTCGTGGCG	660
130470	DB	GCAGTTCACGCTACTTCCTGTGGGTGGGCTCAGTAGAGACTGGATCCGTGTTCGTGGCG	130411
661	QY	CGCAGCGCTTTGTCTCTATCTTACATGATGTTCGCGCTGGGGTGCTACGTGTTCAATCAT	720
130410	DB	CGCAGCGCTTTGTCTCTATCTTACATGATGTTCGCGCTGGGGTGCTACGTGTTCAATCAT	130351
721	QY	CAACCTCATACCCCTGCACGCTCTTCGTCTGTCTCATATGCGCAGGTACTCGCCGCGTCT	780
130350	DB	CAACCTCATACCCCTGCACGCTCTTCGTCTGTCTCATATGCGCAGGTACTCGCCGCGTCT	130291
781	QY	GCTGACGAGCTACGACGACCTTCTACATCTCTGGGACTGTCTGTTCCTCATGCGAGTCCGCTT	840
130290	DB	GCTGACGAGCTACGACGACCTTCTACATCTCTGGGACTGTCTGTTCCTCATGCGAGTCCGCTT	130231
841	QY	CGTGGGATTCGAACCGATACGACAGTGAACACATGCTGCGCTGGGAGTGTGTGTGCT	900
130230	DB	CGTGGGATTCGAACCGATACGACAGTGAACACATGCTGCGCTGGGAGTGTGTGTGCT	130171
901	QY	CCCTATGGCGTGGCGACCTTGGCCCATTTGCAGTCCGTGCTGCGCAACAGATTTCCG	960
130170	DB	CCCTATGGCGTGGCGACCTTGGCCCATTTGCAGTCCGTGCTGCGCAACAGATTTCCG	130111
961	QY	GAAGCTGTTCATCGTGGGGGATTGCTGCTGGCGCTGCGCGTCTTTGTGGCGCTGCTGGT	1020
130110	DB	GAAGCTGTTCATCGTGGCGGATTGCTGCTGGCGCTGCGCGTCTTTGTGGCGCTGCTGGT	130051
1021	QY	GCTCACATCTGGGGTGTGGCCCGGTGGAGTGGAGCGCTTCTACTCGCTGTGGGATAC	1080
130050	DB	GCTCACATCTGGGGTGTGGCCCGGTGGAGTGGAGCGCTTCTACTCGCTGTGGGATAC	129991
1081	QY	TGGCTAGCCCAAGATCCAACTCCCATCATTCGATCCGCTGTGGAGCATCAGCCCAACAC	1140
129990	DB	TGGCTAGCCCAAGATCCAACTCCCATCATTCGATCCGCTGTGGAGCATCAGCCCAACAC	129931
1141	QY	TTGGTTCCTCTTCTTTGATCTGCACATCTGTGTGGCGCTTCCAGTGGGAGTGTG	1200
129930	DB	TTGGTTCCTCTTCTTTGATCTGCACATCTGTGTGGCGCTTCCAGTGGGAGTGTG	129871
1201	QY	GTACTGCATCAAGCAGATCAACGACGCGCGCTTTTCGTGTGTGTGTGTGTGTGTGTGTG	1260
129870	DB	GTACTGCATCAAGCAGATCAACGACGCGCGCTTTTCGTGTGTGTGTGTGTGTGTGTGTG	129811
1261	QY	GGTTTACTTCGCTGGTGTGATGGTGGCTTTGATGTGACCTCAGCCGCGTGTGTGAT	1320
129810	DB	GGTTTACTTCGCTGGTGTGATGGTGGCTTTGATGTGACCTCAGCCGCGTGTGTGAT	129751
1321	QY	GCTGGCGGAGTGGCTTTTCGGGACTGTGTGATGTGTTCCTGCAAGAGGATTCGTCTAA	1380
129750	DB	GCTGGCGGAGTGGCTTTTCGGGACTGTGTGATGTGTTCCTGCAAGAGGATTCGTCTAA	129691
1381	QY	GCNAATGGGCACAGCCATTAAGCGACCCACCGAAGTGATGNACTGGAGTCCCATGA	1440
129690	DB	GCNAATGGGCACAGCCATTAAGCGACCCACCGAAGTGATGNACTGGAGTCCCATGA	129631
1441	QY	GAAGAAGACGCTATACGA-----	1460
129630	DB	GAAGAAGACGCTATACGA-----	129571

QY 1461 -----AGGTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG 1501
Db 129570 TAAATACATCGCATTTTCAGCGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG 129511
QY 1502 CAGGATCTGGCGTTCAGCTCAACCTGAAGAGTATGTTATTTGGCGGTTCTAATGCTG 1561
Db 129510 CAGGATCTGGCGTTCAGCTCAACCTGAAGAGTATGTTATTTGGCGGTTCTAATGCTG 129451
QY 1562 TTGATGATGTTGCGTTCAGCTCAACCTGAAGAGTATGTTATTTGGCGGTTCTAATGCTG 1621
Db 129450 TTGATGATGTTGCGTTCAGCTCAACCTGAAGAGTATGTTATTTGGCGGTTCTAATGCTG 129391
QY 1622 ATTGCTTGGCTTTCCACCAAGCTCAAGATG----- 1652
Db 129390 ATTGCTTGGCTTTCCACCAAGCTCAAGATG----- 129331
QY 1653 -----GATCCGCAACATTTTGAAGAGTATGTTATTTGGCGGTTCTAATGCTG 1684
Db 129330 TTTAAACCTATTAACCTTTTCTCCACTAGATCCGCAACATTTTGAAGAGTATGTTATTTGGCGGTTCTAATGCTG 129271
QY 1685 GCTTACTACTGCTTTTCAGCAACACTCCGATGATGCTCGGTTATGTTCTTGGTGGAT 1744
Db 129270 GCTTACTACTGCTTTTCAGCAACACTCCGATGATGCTCGGTTATGTTCTTGGTGGAT 129211
QY 1745 TACGATACCATAGAGCGGAAATGGCAACAGACGCTAGTGGATAATAATGCTGG 1804
Db 129210 TACGATACCATAGAGCGGAAATGGCAACAGACGCTAGTGGATAATAATGCTGG 129151
QY 1805 AACATAGTCATAGAGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTAC 1864
Db 129150 AACATAGTCATAGAGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTAC 129091
QY 1865 GAAATATGACATCTCTGAGTGGACTACGTTTGGTGTATCTTTGGCGGTTGATGGC 1924
Db 129090 GAAATATGACATCTCTGAGTGGACTACGTTTGGTGTATCTTTGGCGGTTGATGGC 129031
QY 1925 TATCTGCGATGATCAACAGTTCCTGTTGGATGTTCCGAAATGCTGAGGAGAGCAT 1984
Db 129030 TATCTGCGATGATCAACAGTTCCTGTTGGATGTTCCGAAATGCTGAGGAGAGCAT 128971
QY 1985 CCCAAGGACATTAAGGAAGCGATTAATTTCCAGACCGCGGTGAATTCAGGTTAGTGC 2044
Db 128970 CCCAAGGACATTAAGGAAGCGATTAATTTCCAGACCGCGGTGAATTCAGGTTAGTGC 128911
QY 2045 GAAGTGTCTCCGCGCTCTCACTGCTTATGATGACAAATTAAGCTACTACAGATCGG 2104
Db 128910 GAAGTGTCTCCGCGCTCTCACTGCTTATGATGACAAATTAAGCTACTACAGATCGG 128851
QY 2105 GAATTGAAGTTGGACTA----- 2121
Db 128850 GAATTGAAGTTGGACTAAGGTAAGCGAAACATTTCTCTCAGGTAGCGATGACACTMA 128791
QY 2122 -----CAGAGGTCCATCTGGATATGATCGCACAGCTAAGCGCGTCAATTTGGGAATA 2171
Db 128790 CTGCTCTGTTTGAAGGTTCATCTGGATATGATCGCACAGCTAAGCGCGTCAATTTGGGAATA 128731
QY 2172 AGGACTTCGATCTGACCTACTGAGGAGGCTTACACACAGAACACTGCTTTGTTGCA 2231
Db 128730 AGGACTTCGATCTGACCTACTGAGGAGGCTTACACACAGAACACTGCTTTGTTGCA 128671
QY 2232 TCTATAGGTTGAAGCGGATGATTTCAATAGACCATCACTGAAGCAAGAGAGAA 2291
Db 128670 TCTATAGGTTGAAGCGGATGATTTCAATAGACCATCACTGAAGCAAGAGAGAA 128611
QY 2292 CGATTCTCCAGCAAACTTCAATTCGAGAA----- 2321
Db 128610 CGATTCTCCAGCAAACTTCAATTCGAGAAAGGTAGTGTACCTTGCAGCTCCCAATGAAC 128551
QY 2322 -----AGAACTCTAGCGTCGCAAGGCTACATCGAAACC 2357
Db 128550 TTCATTAACAATCTGATTTTTCAGAACTCTAGCGTCGCAAGGCTACATCGAAACC 128491
QY 2358 GACCGGTTGTTTGAAGGGAACCAAGACCTTTGAAATAAACCACCAAAAAA 2412

Db 128490 GACCGGTTGTTTGAAGGGAACCAAGACCTTTGAATAAACCAAAAAAATAA 128436
RESULT 8
LOCUS AC018145/c
DEFINITION Drosophila melanogaster, 75650 bp DNA linear HTG 09-DEC-1999
AC018145
ACCESSION AC018145
VERSION 1
KEYWORDS HTG; HTGS PHASE2
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
ADAMS.M. and Venter, J.C.
DIRECT SUBMISSION
SUBMITTED (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10213542 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
1..75650
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match 88.2%; Score 2131.6; DB 2; Length 75650;
Best Local Similarity 91.4%; Pred. No. 0;
Matches 2407; Conservative 0; Mismatches 4; Indels 223; Gaps 5;
QY 1 TCTAAGCGAAGATGTCGTTGCAATTCAGATCGGTATTAATTTTCGAGTTACTGCTG 60
Db 21557 TCTAAGCGAAGATGTCGTTGCAATTCAGATCGGTATTAATTTTCGAGTTACTGCTG 21498
QY 61 GAAATGGGACATGAATCGGACCGCAAGATGCTCAACAGCAAGTGGCTGGCTACAGCAG 120
Db 21497 GAAATGGGACATGAATCGGACCGCAAGATGCTCAACAGCAAGTGGCTGGCTACAGCAG 21438
QY 121 CTTAATCACTTCGCGATTCCTGTAATCGCTGGCTGGCGGATTTTCCTCTGCTCTT 180
Db 21437 CTTAATCACTTCGCGATTCCTGTAATCGCTGGCTGGCGGATTTTCCTCTGCTCTT 21378
QY 181 CGCGGTATCCGTTTCGAGTCGATTCATTCATGATTTGATCCGTTCAACTACCGGC 240
Db 21377 CGCGGTATCCGTTTCGAGTCGATTCATTCATGATTTGATCCGTTCAACTACCGGC 21318
QY 241 CACCGCTTACATGTTGCGAATGTTGGTACAACTTCCTCAACTGGTTGACAGCGCGC 300
Db 21317 CACCGCTTACATGTTGCGAATGTTGGTACAACTTCCTCAACTGGTTGACAGCGCGC 21258
QY 301 ATGTTATCCGTTTCGCGAGGATTTGGCGGCTACCGTCTATCCCGGCTGATGATTACGTC 360
Db 21257 ATGTTATCCGTTTCGCGAGGATTTGGCGGCTACCGTCTATCCCGGCTGATGATTACGTC 21198
QY 361 CGCGGAAATCACTTGGCTGTCAGTCTCAACATACCGGTCATATTCGTCGATCTG 420
Db 21197 CGCGGAAATCACTTGGCTGTCAGTCTCAACATACCGGTCATATTCGTCGATCTG 21138
QY 421 CGTGTTCCTGCGCGGATTCCTAGTGGCTGACCTCTCACTCTACCTGCTGACCA 480
Db 21137 CGTGTTCCTGCGCGGATTCCTAGTGGCTGACCTCTCACTCTACCTGCTGACCA 21078
QY 481 GGAGCTGTGTTCCGCGGCGCGCTCTTCGCGCGCGCTTCATCGCATCTGCTG 540
Db 21077 GGAGCTGTGTTCCGCGGCGCGCTCTTCGCGCGCGCTTCATCGCATCTGCTG 21018


```

IMAGE:6837097), complete cds.
BC052433
VERSION BC052433.1 GI:30851501
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4236)
Strausberg,R.B., Feingold,B.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heien,F.,
Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldi,M.F., Casavant,T.L.,
Schett,T.B., Brownstein,M.J., Uedin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Pahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalek,U., Smalish,D.E.,
Schnerch,A., Schein,J.B., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
Strausberg,R.
Direct Submission
Submitted (15-MAY-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcaphs-rc@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas L. Casavant.
Web site: http://genome.uiowa.edu
Contact: bento-soares@uiowa.edu, tom-casavant@uiowa.edu
Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
Fisher,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,
Casavant,T., Soares,M.B.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
1. .4236
/organism="Mus musculus"
/ncbi_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:64679 IMAGE:6837097"
/tissue_type="Brain, enriched mouse brain 12.5dpc"
/clone_lib="NIH BXAP_F10"
/lab_host="DH105"
/note="Vector: pYX-ASC"
1. .4236
/gene="1300006C19Rik"

```

/note="synonym: Simp"
 /db_xref="LocusID:88292"
 /db_xref="MGI:1915542"
 366. .2837
 /gene="1300006C19Rik"
 /codon_start=1
 /product="1300006C19Rik protein"
 /protein_id="AAH52433.1"
 /db_xref="GI:30851502"
 /db_xref="LocusID:88292"
 /db_xref="MGI:1915542"
 /translation="MAEPSAPESKHSLSINSPWGLMAGNRSRHGHGPGTQSASRA
 AAKPFGAGLSQAGWQSLSTFLIPLAWGFSRLFAVIRFVILHEDPDM
 FNRSTHLSHGFEFLNFDERAWYPLGRIVGVTVVGLMITAGLHWLITHT
 VHIRDVFLAPTSGLSISFLTLRELWQAGLLAACFIAIVPGIISRVAGSFD
 NRGIAIFALQTYLWVSKVSGFVWMCCLSFYVWSAWGGVFIILPLHVFV
 LLUMRYKRYIATSYFVGLILSMQIPFQFPIRTSEMAAGVFIALLQAYAF
 QIDRLRLKQEFQTLFFGLVSLAAGVLSVILTYITGYIAPWSGRFYSLMDTGYAKI
 HIPIIASVSEHQPTTWVFFDFDLHLVCTFFAGLWFCIKNINDERFVVALYSAIVF
 AGVMVRLMLTLTPVCMLSAIFSNVFEHLGDDMKRENPPVEDSSDDKRNPNGLV
 DKAGVKRVHTQEKPEGLGNIKSVITMLMLLMVFAVHCTWVTNAYSSPSVL
 ASYNDGTRNILLDDPREAFVWLRLONTDEHSAVMSWDYGYQIAGNRTILVDNNTN
 NSHIALVGKMSNETAAYKMRSLDVLVIFGVGIGSGDDINKELMWRVIAEGE
 HPKDRGDIYFQQGEFVDRKAGSFTLLNCLMYKMSYRFGEOMDLFFTPGFDTRN
 ABIGNKIDFKHLEAFTSEHLVRIYKVPADNRETTLGHRVTNIVPKQKLSKKT
 TKRKGYVKNLVEFKGKTKTV"
 567. .2837
 /gene="1300006C19Rik"
 /note="STT3: Region: Oligosaccharyl transferase STT3
 subunit. This family consists of the oligosaccharyl
 transferase STT3 subunit and related proteins. The STT3
 subunit is part of the oligosaccharyl transferase (OTase)
 complex of proteins and is required for its activity.
 OTase transfers a lipid-linked core-oligosaccharide to
 selected asparagine residues in the ER"
 /db_xref="CDD:pfam02516"

Query Match 39.8%; Score 962.2; DB 10; Length 4236;
 Best Local Similarity 64.8%; Pred. No. 6.8e-207;
 Matches 1496; Conservative 0; Mismatches 793; Indels 18; Gaps 4;
 QY 102 AGTGGCTGGCTACAGCAGCCTAATCAGCTCGCCATCCTCTAATCGCTGGCTGGCGG 161
 DB 547 AGCGCGCGGTGGAGCTGTTGCTCTCCITTCACATCCTCTCTCGCTGGCTGGCGG 606
 QY 162 GATTTCCTCGCTCTTCGCGCTCATCGCTTCGAGTCGATTCATTCATGAGTTTCATC 221
 DB 607 GCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAGAGCATCATCCAGAGTTCCACC 666
 QY 222 CGTGGTTCAACTACCGGCGCACCGCTCATGTCGAGATGTTGGTGAACCTTCCTCA 281
 DB 667 CGTGGTTTAACTAGATCAACACATCATCTTCGATCTCATGGATTCATGAGTTTCTAA 726
 QY 282 ACTGGTTTCGAGAGCGGCGCATGTTATCGCTCGGCGGATTTGGGCGGTACCGTCTATC 341
 DB 727 ATTGGTTTGAAGAAGAGCATGTTACCACTGGGAAGATAGTGGTGGCACCGTTTACC 786
 QY 342 CCGGCTCTGATGATGATCGTCGGCGGAATCCATTGGCTGTCGACGTACTCAACATACCGG 401
 DB 787 CAGGTTTGAATGATTAACAGCTGGCTTATTCATTGGATTTTAAATACATTGAACATAACAG 846
 QY 402 TCCATATTCGTGACATCTCGTGTTCCTCGCGCGGATCTTCAGTGGCTGACCTCCATCT 461
 DB 847 TTACATAAGAGATGTGTGTATTCCTTGACACAACTTTAGCGGCTTACATCCATAT 906
 QY 462 CCACCTACTGTGTACCAAGAGCTGTGTCGCGGCGCGCGGCTCTTCGCGCGGAGCT 521
 DB 907 CTAGTTCCTGCTAACTAGAGAACTGTGGAACCAAGAGGAGGACCTTCTAGCTGCTTCT 966
 QY 522 TCATCGGCATCGTGGCTGCTACATGATGAGTGGTGGCTGGCTGATCGTAGATAACAGG 581
 DB 967 TCATGCTATCGTACAGGCTGATATCTCGGTGAGTGGCGGATCTTTTGTGATTAATGAAG 1026

ORIGIN
 Query Match 39.8%; Score 962.2; DB 10; Length 4236;
 Best Local Similarity 64.8%; Pred. No. 6.8e-207;
 Matches 1496; Conservative 0; Mismatches 793; Indels 18; Gaps 4;
 QY 102 AGTGGCTGGCTACAGCAGCCTAATCAGCTCGCCATCCTCTAATCGCTGGCTGGCGG 161
 DB 547 AGCGCGCGGTGGAGCTGTTGCTCTCCITTCACATCCTCTCTCGCTGGCTGGCGG 606
 QY 162 GATTTCCTCGCTCTTCGCGCTCATCGCTTCGAGTCGATTCATTCATGAGTTTCATC 221
 DB 607 GCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAGAGCATCATCCAGAGTTCCACC 666
 QY 222 CGTGGTTCAACTACCGGCGCACCGCTCATGTCGAGATGTTGGTGAACCTTCCTCA 281
 DB 667 CGTGGTTTAACTAGATCAACACATCATCTTCGATCTCATGGATTCATGAGTTTCTAA 726
 QY 282 ACTGGTTTCGAGAGCGGCGCATGTTATCGCTCGGCGGATTTGGGCGGTACCGTCTATC 341
 DB 727 ATTGGTTTGAAGAAGAGCATGTTACCACTGGGAAGATAGTGGTGGCACCGTTTACC 786
 QY 342 CCGGCTCTGATGATGATCGTCGGCGGAATCCATTGGCTGTCGACGTACTCAACATACCGG 401
 DB 787 CAGGTTTGAATGATTAACAGCTGGCTTATTCATTGGATTTTAAATACATTGAACATAACAG 846
 QY 402 TCCATATTCGTGACATCTCGTGTTCCTCGCGCGGATCTTCAGTGGCTGACCTCCATCT 461
 DB 847 TTACATAAGAGATGTGTGTATTCCTTGACACAACTTTAGCGGCTTACATCCATAT 906
 QY 462 CCACCTACTGTGTACCAAGAGCTGTGTCGCGGCGCGCGGCTCTTCGCGCGGAGCT 521
 DB 907 CTAGTTCCTGCTAACTAGAGAACTGTGGAACCAAGAGGAGGACCTTCTAGCTGCTTCT 966
 QY 522 TCATCGGCATCGTGGCTGCTACATGATGAGTGGTGGCTGGCTGATCGTAGATAACAGG 581
 DB 967 TCATGCTATCGTACAGGCTGATATCTCGGTGAGTGGCGGATCTTTTGTGATTAATGAAG 1026

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REMARK
 COMMENT

NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk
 Email: gcaphs-rc@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
 Thomas L. Casavant.
 Web site: http://genome.uiowa.edu
 Contact: bento-soares@uiowa.edu, tom-casavant@uiowa.edu
 Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
 Fisher,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
 Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,
 Casavant,T., Soares,M.B.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
 Series: Plate: Row: Column: 0
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Similarity but not identity to protein.

Location/Qualifiers
 1. .4236
 /organism="Mus musculus"
 /ncbi_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="MGC:64679 IMAGE:6837097"
 /tissue_type="Brain, enriched mouse brain 12.5dpc"
 /clone_lib="NIH BXAP_F10"
 /lab_host="DH105"
 /note="Vector: pYX-ASC"
 1. .4236
 /gene="1300006C19Rik"

gene

/mol_type="mrna"
/db_xref="taxon:10090"

ORIGIN

Query Match 39.7%; Score 960.6; DB 6; Length 2710;
Best Local Similarity 64.8%; Pred. No. 1.5e-206;
Matches 1495; Conservative 0; Mismatches 794; Indels 18; Gaps 4;
102 AGTGGCTGGCTACAGCAGCCTAATCACTTGGCCATCTCTGCTAATCGCTGCTGCGCG 161
Db AGCGGCGCGGTGGCAGTGGTGTCTCTCTTCAACATCTCTTCTGCGCTGCGCGCG 317
162 GAATTTCTCTCTGCT 221
Db GCTTCAGCTCTGCGCT 377
222 CGTGGTTCAATACCGGCGCCACCGCTACATGGTGCAGAAATGGTGGTGAACAATCTCTCA 281
Db CGTGGTTTAACTATAGATCAACACATCATCTCTGATCTCTGATCTCTGATCTCTGATCT 437
282 ACTGGTTTCAGAGCGCGCATGGTATCGCTCGGCGAGATTTGGGCGGTACCGTCTATC 341
Db ATTGGTTTGAAGAGAGATGGTACCCACTTGGGAGAAATAGTGGTGGCACCGTTTACC 497
342 CCGGCTGATGATACGTTCGGCGGAATCCATTTGGCTGCTGCAGATCTCAACATACCGG 401
Db CAGGGTTGATGATAACAGCTGGCGCTTATTCATTGGATTTTAAATACATTTGAACATACAG 557
402 TCCATATTCGTGACATCTGGCTGTCTCTGGCGCGCATCTTCAGTGGCTCACCTCCATCT 461
Db TTACATAAGAGATGTGTGTATCTCTTGACCAACTTTTAGCGGCTTTACATCCATAT 617
462 CCACCTACTCTGTACCAAGAGAGTGTGGTCCGGCGCGCGCTCTTCGCGCGCAGCT 521
Db CTACCTCTCTGTAACCTAGAGAACTGTGGAACTTGAATTTTAAATACATTTGAACATACAG 677
522 TCATCGCATCTGCTGCTGTACATCAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 581
Db TCAATTTGCTATCGTACAGAGTACATATCTCGGTGAGTGGCGGATCTTTTGAATATGAAG 737
582 GAATTTGCAATATCGCTCGAGTTTACCTTCTCTGTGGTGGTGGTGGTGGTGGTGGTGG 641
Db GCAATTTGCCATTTTTCGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 797
642 GATCGGTGTTCTGGTGGCGCGCGCTTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 701
Db GGTCTGTGTTCTGGACATGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 857
702 GTGGTACGTGTTCATCAACCTGTATACCCCTGCACTCTCTCTCTCTCTCTCTCTCTCT 761
Db GAGGTATGTGTTTCACTCAACCTCACTCACTCACTCACTCACTCACTCACTCACTCACT 917
762 GCAGTACTCGCGCGCTGCTGTGACCACTACAGCTCTCTCTCTCTCTCTCTCTCTCTCT 821
Db AGAGGTACAGCAGAGAGTCTACATAGATATAGATCTTGTATCTTGTGGTGTATATAT 977
822 TCTCATGAGATATCCCTTGTGGATTCGAACCGATAGCAGCAGGATGAAACAATGGCTG 881
Db TATCATGAGATATCCCTTGTGGATTCGAACCGATAGCAGCAGGATGAAACAATGGCTG 1037
882 CGCTGGAGTGTCTGTGCTCTTATGGCGGTGGCGACCTTGGCGCATTTCCAGTCCGTGC 941
Db CTGCAAGTGTCTTGGCTGCTGCAAGCTTACGCTTTTGTGATCTCTGAGAGACGGT 1097
942 TGTGCGGCAACAGATTCGGAGAGTGTTCATCTGCTGGCGGATTTGCTGGTGGCGGTGGCG 1001
Db TGACAAAACAGAGATTCGAGACCTCTTTCTTTTGGGTGTCTCACTAGCTGCGAGCGCTG 1157
1002 TCTTTGTGCGCTGTGTGTCTCACTATGCTGGCGTGTGGCGCTGTGGCGCTGTGGAGCTG 1061
Db TGTTCTTGT 1217
1062 TCTACTCGCTGGGATATCTGGCTACGCCAAGATTCACATTCCTCATTTGCCATTCCTGTG 1121

1218 TTTATTCATATGGGATATCTGGGTATGCAAAAAATACACATTCCTCAATTTATGATCATGT 1277
QY CGGAGCATCAGCCACCATTTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1181
Db CTGAACATCAGCTACGACATGGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1337
1182 CTTTCCAGTGGGAGTGTGTATCTGATCAAGCAGATCAACGAGAGCGGTTTTCGTGG 1241
Db CTTTCCAGTGGGAGTGTGTATCTGATCAAAAAATATCAACGATGAAGAGATTTTGTGG 1397
1242 TGCTGTACGCGCATCAGTGGCGTCTTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1301
QY CTCTGTATGCGATCAGTGTCTGTACTTTCGCGAGTGTGTGTGGCTGAGTGTGACTC 1457
1302 TCAGCGCGGTGTGTGTATCTGCTGGCGGAGTGGCTTTTCGGGACTGTGTGGATGTGTCTC 1361
Db TGACCGCGGTGTGTGTATCTGCTGGCGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1517
1362 TGCAAGAGGATTCGTCTAAGCGAATGGGACAGCCATTAAGCGCAGCCACCGAAGTGGATG 1421
Db TGGGGATGA-----CATGAAAGGGAAACCCACCTGTGGAG-----GACAGCAGTG 1565
1422 AAGCTGAGGATTCATTTGAGAAAGAGCGCTGTACGAAAGCGCTGCGAGCTGGAAGCTGA 1481
Db ATGAGGATGACAAAGAAACCCAGGAACTTGTATCAAGCGCAGGTAAAGTGAAGAGC 1625
1482 GTACTAAGCATGTATGCCAGCAGGATCTGGCGTCTGAGCTTCAACCTGAAAGATTTATGTTA 1541
Db ATGTGACAGAGCAAGAAACCTGAAGAGGCTTGGGCCCAACATCAAAAGCATTTGTGA 1685
1542 TTTTGGCGGTTCATCTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1601
Db CCATGCTGATGCTCATCTCTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1745
1602 ATGCTACTCTCAGTCCCTCCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1661
Db ACGCTACTCTCAGTCCCAAGT 1802
1662 ACATTTAGAGATTCAGAGAGCTTACTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1721
Db ATATATTAGATGATTTTAGAGAGCGTACTTTTGGTGTGACAAACACCGGATGAACAGC 1862
1722 CTCCGTTATGTCTTGTGTGGATTTAGGATACAGATAGCGGAAATGGCAACAGAAACA 1781
Db CCGGCTCATGTCTGTGTGGATCTAGGCTATCAGATTTGCTGTGCTGCGCCACAGGACCA 1922
1782 CGCTAGTGGATTAATAATACCTGGAAACAATAGTCAATAGCGCTGTGTGTGTGTGTGTGT 1841
Db CTCTGTGGATTAACAACCTGGAAACAACAGCCACATCGCATCTGTGTGTGTGTGTGTGTGT 1982
1842 CTTCAACCGAGGAGAGTCTCTAGAAATATGACATCTCTGACGTGGACTACGTTTGTGG 1901
Db CTTCAATGAAACCGCGCCCTATATAATCATGAGTCCCTGTGATGTGATGTGTGTGG 2042
1902 TGATCTTTGGCGGT 1961
Db TTTTCTGGAGGATGATTTGGCTATTCGCGGGAACGATATCAACAAAGTTCTCTGTGTGG 2102
1962 TCCGAATTTGTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGGATTTACTTTACCGACC 2021
Db TCAGATAGCTGAAGGGGAGCATCCCAAGACATCCGGAAGGTGACTATTTCCACCAGC 2162
2022 CGGTGTAAATTCAGGGTATAGTGGAGGTGCTCCGCGCTGTCTCACTCTCTCTCTCTCTCT 2081
Db AGGAGAGATTCAGGATGAGCAAAAGCTGGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2222
2082 AATTAAGCTACTACAGATTCGCGGATTTGAAGTTGACTACAGAGTCCATCTGGAATG 2141
Db AATGTCTACTACAGATTTGGAGAAATGCAGCTAGATTTTCGACTCTCCCGCAGCTTTG 2282
2142 ATCGCACAGCTAACCGCGTCAATTTGGGAATGAAGACTTCGATCTGACCTTACCTTGGAGG 2201

Db 2283 ACCGACACGTAATGCTGAGATTGGAATAAAGACATTAAATTCAGCAATTTGGAGGAAG 2342

QY 2202 CCTACACCAGAACACTGGCTGTTGCGCATCTATAGGTTGAAGCGCGCATGAGTTCA 2261

Db 2343 CTTTATCATCAGAGCAGCTGGCTTGTGAGGATATATAAGTGAAGACCTGACACAGG 2402

QY 2262 ATAGACCATCACTGAAGACCAAGAGAGAGAACGATTCTCCAGCAAACTTCATTTCAGAA 2321

Db 2403 AGACACTAGGTCAACAACTCGAGTCAACCAACATCGTCCCAACACAGAGATTGTCAA 2462

QY 2322 AGAAGCTTAAAGCTGCGAAG--GGGTACATACAGAACCGACCGCTTGTGTAAGGAA 2378

Db 2463 AGAAGACTTAAAGAGAGCGTGGCTAGCTTAAATAAGCTAGTGTGTAAGAAAGCA 2522

QY 2379 AACGAACCTTGAATAAACCACAAAAA 2405

Db 2523 AGAAGACCTCTAAGAGACTGTTTAAA 2549

RESULT 11

AX799082 AX799082 2481 bp mRNA linear PAT 08-OCT-2003

LOCUS Sequence 1 from Patent WO03054008.

DEFINITION AX799082

ACCESSION AX799082

VERSION AX799082.1 GI:37605057

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE Perreault, C. and McBride, K.

JOURNAL Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy

FEATURES

Location/Qualifiers

1..2481

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

1..2481

/codon_start=1

/product="SIMP"

/protein_id="CAB48447.1"

/db_xref="GI:37605058"

/translation="MAEPGAPESKHKSLSNPSWGLMALGNSRHGHGPGAAQAHKA AGGAAPPKAPAGLSQAPAGWQSLLSFTILFLAWLAFSPSSLLFAVIRPESIIHEF DPWNVKSTHLLASHGFYEFPLNWFDERAWYPLGRIVGGTVYPLGLITAGLILHILNTL NITVHIDVCVFLAPTFFSLTISITFLPRELWNOGAGLLAACPIAIVPGVYSRVAG SPNEGIAIFALQFTYLVKSVKTSQVFWNTMCCLSYFVMSVANGVVELINLPIH VFVLLLMQRYSKRYIAYSTFIIVGLILSMQIPFVPGPDIPTSEHMAAGVFPALLOAY AFLOQRDUTKQFEFTFLFPLGSLAAGAVFLSVIYLTITGTIAPWSGRFVSLMDTQY AKIHPILIASVSHQPTTWVSFFDLHLVCTFFPAGLWFCIKNINDERFVALYISA VYFAGVVRMLTLTTEVCMLSAIAFSPNFEHLGDDMKRENPPVDESSDDKRNQG NLVDKAKYRKHATEQEBTEGLGNPKISITVLMMLMLMFVAVCTWNTSNAYSSPS VLVASYNHDTIRNILDREAFYFWLRQNTDEHARVMSWMDYQIAGMANTTILVDNN TWNSHIALVKGKMSNTAAKYIMETLDDVVLVIFGVIGVSGDDINKFLMWVRIA EGHKHDILRESDFTPQGRFPKDGKSPILLNCLMYKSVYRFGEMOLDPETPGPR TRAEIKNKDIKEKHEAFTSEHMLVRIYKYKAPDNRETLDRPVRNIPFKKYLIS KKTTRKRGYIKKLVFKKXKISIKTV"

CDS

1..2481

/product="SIMP"

/protein_id="CAB48447.1"

/db_xref="GI:37605058"

/translation="MAEPGAPESKHKSLSNPSWGLMALGNSRHGHGPGAAQAHKA AGGAAPPKAPAGLSQAPAGWQSLLSFTILFLAWLAFSPSSLLFAVIRPESIIHEF DPWNVKSTHLLASHGFYEFPLNWFDERAWYPLGRIVGGTVYPLGLITAGLILHILNTL NITVHIDVCVFLAPTFFSLTISITFLPRELWNOGAGLLAACPIAIVPGVYSRVAG SPNEGIAIFALQFTYLVKSVKTSQVFWNTMCCLSYFVMSVANGVVELINLPIH VFVLLLMQRYSKRYIAYSTFIIVGLILSMQIPFVPGPDIPTSEHMAAGVFPALLOAY AFLOQRDUTKQFEFTFLFPLGSLAAGAVFLSVIYLTITGTIAPWSGRFVSLMDTQY AKIHPILIASVSHQPTTWVSFFDLHLVCTFFPAGLWFCIKNINDERFVALYISA VYFAGVVRMLTLTTEVCMLSAIAFSPNFEHLGDDMKRENPPVDESSDDKRNQG NLVDKAKYRKHATEQEBTEGLGNPKISITVLMMLMLMFVAVCTWNTSNAYSSPS VLVASYNHDTIRNILDREAFYFWLRQNTDEHARVMSWMDYQIAGMANTTILVDNN TWNSHIALVKGKMSNTAAKYIMETLDDVVLVIFGVIGVSGDDINKFLMWVRIA EGHKHDILRESDFTPQGRFPKDGKSPILLNCLMYKSVYRFGEMOLDPETPGPR TRAEIKNKDIKEKHEAFTSEHMLVRIYKYKAPDNRETLDRPVRNIPFKKYLIS KKTTRKRGYIKKLVFKKXKISIKTV"

ORIGIN

Query Match 39.5%; Score 955.8; DB 6; Length 2481;

Best Local Similarity 64.8%; Pred. No. 1.8e-205;

Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;

QY 102 AGGTGGCTGGCTACAGCAGCTTAATCACCTTCGCCATCGCTTAATCGCTGGCTGGCG 161

Db 191 AGCCGGCTGGGTGGAGTCGCTTCTCTCTTACCATCTCTTCTGGCTGGCTGGCG 250

QY 162 GATTTCCTCTCGCCTCTTCGCGCTCATCGGTTTCGAGTCGATTATCCATGAGTTGATC 221

Db 251 GCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAAAGCATCATCCAGAGTTGGACC 310

QY 222 CQTGTTCACTACCGGCGCACCGCTTACATGTTGTCAGATGTTGTTGTAACAATCTCTCA 281

Db 311 CQTGTTTAACTATATAGATCAACATCATCTTTCATCTCAITGGGTTCTATGAATTTTAA 370

QY 282 ACTGGTTCCAGAGCGCGCATGTTATCGCTCGCAGAGATTGTGGCGGTACCGTCTATC 341

Db 371 ATTGGTTTATGAAAGAGCATGGTATCCACTTAGGAAGATAGTAGGTGTACTGTTTACC 430

QY 342 CCGGCTGATGATTACGTCGCGGGAATCCATTCGCTGCTGCAGTACTCAACATACCGG 401

Db 431 CAGGGTTGATGATAACCGCTGGCTTATTCATTCGATTTTAAATACATTCGAAATACATG 490

QY 402 TCCATATTGTTGACATCTCGCTGTTCTCGGCGCGATCTTCAGTGGCTCACTCATCT 461

Db 491 TCCATTAAGAGACGATATGTTGTTCTTCGACCAACTTTTAGCGGCTTTACATCTATAT 550

QY 462 CCACCTTACCTGCTGACCAAGAGAGCTGTTCCGCGGCGCGGCTCTTCGCGCGCAGCT 521

Db 551 CTACTTCTCTGTTTAAAGAACTTTGGAAACCAAGAGAGAGACTTTTACGTTGTTT 610

QY 522 TCATCGCCATCGTCTCGCTTACATCAGTAGGTTCGTTGGTGGATCGTACGATAACGAGG 581

Db 611 TTATTGCTATTGTACAGGCTACATATCTCGTTCAGTAGCTGGATCTTTGATAAATGAAG 670

QY 582 GCATTGGCATATTCGCGCTCGAGTTCACCTACTTCTTCGTTGGTGGCTCGTCAAGACTG 641

Db 671 GCATTGCTATTTTGCATCTCAGTTCACATCTATTTATGGGTAAATCTGTAAAACCTG 730

QY 642 GATCGCTGTTCTGTCGCGCGCAGCGCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTG 701

Db 731 GGTCAAGTTTGTGCAATGCTGCTGCTTATCTTCTTCTTCTTCTTCTTCTTCTTCTG 790

QY 702 GTGGCTAGGTTCATCATCACTGATACCTGACCTGACCTCTTCGACTCTCTCATTTATG 761

Db 791 GTGGTTATGTTTATCATCACTTATTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 850

QY 762 GCAGGTACTTCGCGCGCTGCTGACAGCTACAGACCTTCTTCTTCTTCTTCTTCTTCTG 821

Db 851 AGAGATACAGCAAAAGAGCTTACATAGCATATAGCACTTTCTTCTTCTTCTTCTTCT 910

QY 822 TCTCATGACATCCCTTCGTTGGGATTCACACCGATACGACAGTGAACATCGCTG 881

Db 911 TATCAATGAGATACCTTTTGGGATTCAGCCATCAGAACAGTGAACATCGCAG 970

QY 882 CCGTGGGAGTGTGCTTCTTATGCGCGTGGCCACCTTCGCGCATTTTCAGTCCGCTGC 941

Db 971 CTGCAAGTGTCTTTGCAATGCTGCAAGCTTATGCTTCTTCTGCACTATCTGAGAGCCG 1030

QY 942 TCTCGCGCAACGAGTTCGCGGAGCTTTCATCGTCGCGGATTCGCTGCGGCTGGCG 1001

Db 1031 TAAACAAACAGAGTTCAGACCTTTCTTTCTTTTGGGTGATCACTAGCTGCGAGTCTG 1090

QY 1002 TCTTTGCGCGCTGCTGCTGCTACCATGCTGGCGTGTGGCGCTTGTGGCGCTGAGTGG 1061

Db 1091 TGTTCCTTAGTGTCTATTTGCTTATACAGTTTACAGTTTACATTCGACCATGCGAGT 1150

QY 1062 TCTACTCGCTGGGATCTGCTGCTAGCGCAAGATCCCATTTCCCATTCATTCGATCCG 1121

Db 1151 TTTATTTCATTTGGGATCTGCGTATGCAAAAATACATTCCTTCTTCTTCTTCTTCT 1210

QY 1122 CCGAGCATCAGCCACCACTTGGTTCTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 1181

Db 1211 CTGAGCATCAACTACGACTTGGGTGTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1270

QY 1182 CTTTCCCTGAGTGGTGTGCTGCTGCTCAAGAGAGATCAACAGAGCGGCTTTCTGTTG 1241

Db 1271 CTTTCCAGAGCGCTTTGGTTCGTCATCAAAAATATCAAGATGAAGAGATTTGTTG 1330

QY 1242 TCGTGTAGCCATCAGTCCGTTTCTTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1301

Db 1331 CTCTATATGCAATCAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1390

QY 1302 TCACGCCGCTGCTGTCATCTGCGCGAGTGGCTTTTCGGGACTGCTGGATGTGCTCC 1361
 Db 1391 TCACCTCCAGTCTGCTGATGCTGCTGCAATGCTTTTCAATGTTTTGACACTATT 1450
 QY 1362 TCAAGAGAGATTGCTTAAGCGAATGGGCAAGCATAAGCGCAGCCACCGAAGTGGATG 1421
 Db 1451 TGGGGGATGA-----CATGAAAGGGGAAATCCACCTGTGGAG-----GACAGCAGTG 1498
 QY 1422 AAGCTGAGATTCCATTTCAGAGAAGACGCTGTACGACAAAGCTGGCAAGCTCAAGCATC 1481
 Db 1499 ATGAGGATGACAAAGAAACCAAGGAATTTGTATGATAGGCGAGGTAAAGTGAAGAAC 1558
 QY 1482 GTACTAAGCATGATCCCGCAGCAGGATCTGCGGTGAGTCCAGCTCAAGAGATTTGTTA 1541
 Db 1559 ATGCAACTCAACAGGAAAGAACTGAAGAGGATAGGCCCTTAATATATAAAGCAITTTGCA 1618
 QY 1542 TTTTGGCGCTTCTAATGCTGTTGATGTTGCTGCTCCACTGCACTGGGTGACACGCA 1601
 Db 1619 CCATGTTAGTGTGATGATGATGATGTTGCTGCTCCACTGTACCTGGGTCAAGCA 1678
 QY 1602 ATGCTACTCCAGTCCCTCCATTTGCTTTGGCTTCCACACAGTCAAGATGGATCCCGCA 1661
 Db 1679 ATGCTACTCTAGTCCAAAGTGTAGTCTGCTGGCTCATACAA---TCATGATGGCACCAGGA 1735
 QY 1662 ACATTTAGCATTTACAGAGGCTTACTACTGCTTTCGCAAGCACTGCCGATGATG 1721
 Db 1736 ATATCTTAGATGATTTTAGAGAGCTTACTTTGGCTTAAGCAAAATACAGATGAACATG 1795
 QY 1722 CTCGCGTTATGCTTTGGTGGATACCGATACAGATAGCGGGAAATGGCAACAGAACGA 1781
 Db 1796 CACGAGTATGCTTTGGTGGATATGCTTATGCTATCAGATAGTGGAAATGGCTTAATAGAACA 1855
 QY 1782 CGCTAGTGATAATAATACGTGGCAACATAGTCAATAGCGCTGGTTGGCAAGCAATGT 1841
 Db 1856 CTTGCTGATATATACACCTGGAATACACCCACATAGACTGCTGGGAAAGCTATGT 1915
 QY 1842 CTTCAACGAGAGAGTCTTACGAAATATGACATCTCTGAGCTGGACTACGTTTGG 1901
 Db 1916 CTTCTAATGAAACAGCAGCCTATAAAATCATGAGGACTCTAGATGTAGATTATGTTGG 1975
 QY 1902 TCATCTTTGGCGGTGATCGCTATCTGCGGATGATATCAACAGTCTCTGCGATGG 1961
 Db 1976 TTATTTTGGAGGGTTATGCTATCTGCTGATGATATCAACAAATTTCTCTGATGG 2035
 QY 1962 TCCGAATTGCTGAGGAGAGATCCCAAGGACATTAAGGAAGCATTAATTACCGAAC 2021
 Db 2036 TTAGGATAGCTGAAGAGAGACATCCCAAGACATTCGGGAAGTGAATTTTACCCAC 2095
 QY 2022 GCGGTGAATTCAGGTAGATCCGAGGTGCTCCGCGCTGCTCAACTGCTTATGTACA 2081
 Db 2096 AGGGAGATTCGCTGATAGCAAGCAGGATCCCTACTTTGTTGAATGCTTATGTATA 2155
 QY 2082 AATTAAGTACTACAGATTCGGGGAAATTAAGTTGGACTACAGAGTCCATCTGGATATG 2141
 Db 2156 AATGCTCATCTACAGATTTGGAGAAATGCACTGGATTTTCGPACACCCCAAGTTTG 2215
 QY 2142 ATCGCACAGTAAACCGCTCTATCGGATAGGACTTCGATCTGACCTACCTCGAGAGG 2201
 Db 2216 ACCGAACAGTAACTGCTGATTTGAATAAGGACATTAATTCACACATTTGGAAGAG 2275
 QY 2202 CTTACACCAAGCACTGGCTTGTTCGCATCTATAGGGTGAAGAACCCCATGATTTCA 2261
 Db 2276 CTTTACATCAGACACTGGCTTGTAGGATATATAAGTAAAGACACCTGATACAGGG 2335
 QY 2262 ATAGACCATCACTGAAGACCAAGGAGAGACGATCTCCAGCAAACTTCATTTCCGAA 2321
 Db 2336 AGACATTAGATCACAACCTCGAGTCACCAACATTTTCCCAACAGAGATTTGTCAA 2395
 QY 2322 AGAATCTAAGCGTCCGAAG---GGCTACATACGAAACCCGACCGGTTGTTGTTAAGGAA 2378
 Db 2396 AGAAGACTACCAAAAGGAGCGTGGCTACATTAATAAATAAGCTGGTTTTTAAGAAGGCA 2455

QY 2379 AACGAACCTTGAAATAAAC 2397
 Db 2456 AGAAATATCTAAGAAGAC 2474
 RESULT 12
 AY074880 2481 bp mRNA linear PRI 10-DEC-2002
 LOCUS
 DEFINITION Homo sapiens source of immunodominant MHC-associated peptides
 (SIMP) mRNA, complete cds.
 ACCESSION AY074880
 VERSION AY074880.1 GI:19879588
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2481)
 AUTHORS McBride K., Baron C., Picard S., Martin S., Boismenu D., Bell A.,
 Bergeron J., and Perreault C.
 TITLE The model Bedomi minor histocompatibility antigen is encoded by a
 mouse homolog of the yeast Slt3 gene
 JOURNAL Immunogenetics 54 (8), 562-569 (2002)
 MEDLINE 22326278
 PUBMED 12439619
 REFERENCE 2 (bases 1 to 2481)
 AUTHORS McBride K. and Perreault C.
 TITLE Direct Submission
 JOURNAL Submitted (22-JAN-2002) Molecular Biology Group, Compatigene, 6100
 Royalmount, Montreal, Qc H4P 2R2, Canada
 FEATURES
 source
 1. .2481
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 1. .2481
 /gene="SIMP"
 1. .2481
 /note="similar to Saccharomyces cerevisiae OGTST3 and Mus
 musculus protein encoded by GenBank Accession Number
 AK018758; source of minor histocompatibility antigen(s)"
 /codon_start=1
 /product="source of immunodominant MHC-associated
 peptides"
 /protein_id="AAL71884.1"
 /db_xref="GI:19879588"
 /translation="MAEPSAPESKHSLSNPSWGLMALGNSRHGHGPGQAQAHKA
 AGGAAPKAPAGLSGCLSGOPAGWQLSFTILFLAWLAGFSRLPAVIRPESIIHEP
 DPWFNVESTHLASHGFEFLNWFDERAVPLGRIVGTVYQLMITAGLIHWLNTL
 NITVHIDVCVFLAPTFSLTSLTRELWNQAGLLAACFTALVPGYISRVAG
 SFNDEGTAIFALQFTYILWVKSVTGVSFWTMCCLSYFYMVSAWGYYVFIINLPLH
 VFVLLMQRYSKRYIAYSTFYIVGLISMQIPVFGQFPIRTSEHMAAGVALLQAY
 AFQYLDRITLQEFQTLFELGSLAAGAVLSVIYLTGYIAPWSAGFYSLMDTGY
 AKHIPTIAGVSEHQTTWTFPFDFLHILVCTFPAGLWFCIKNIINDERVFVLYAIS
 VYPAGVVRMLLTLPVCMLSAIFSNVFEHLGDMKRENPVDESDDEKNOG
 NLXDAKCKKATKTEQEGPNIKSIVTMLMLMLMPFAVHCTTWTSTNAYSPS
 VVLAVNHDCGTNLLDPEAYFLEQNTDEHARVMWMDYGVQIAGMANRTLLVNN
 TWNSHALVCKAMSNETHAYKIMETLDVDYVLVIFCGVIGSGDDINKFLMWRIA
 EGEHPKDIRKSDYTFQGEPRVDKAGSPILLNCIMYKNSIYRFGENQQLDFRTPPGFDR
 TRNAIGNDKIKFKHLEAFTSEHLVIRYIKVAPDNRETLDHKPRVTNIFPKQKYL
 KKTTRKRGYIKKLVFKKIKSKKTV"

ORIGIN

Query Match 39, 58; Score 955.8; DB 9; Length 2481;
 Best Local Similarity 64.8%; Pred. No. 1.8e-205;
 Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;
 QY 102 AGGTGGCTGGGTACAGAGCCTAATCACCTTCGCCATCTGCTAATGCGCTGCGCGC 161
 Db 191 AGCGGCTGGGTGGCAGTGGCTTCTCTCTTCCATCTTCCATCTCTTCTGGCTGGCT 250

Db 2396 AGAAGACTACCAAGAGGCGTGGCTACATTAATAAATAAGCTGGTTTTTAAGAAAGGCA 2455

QY 2379 AACGAACCTTGAATAAATAC 2397

Db 2456 AGAATAATCTAAGAGAGAC 2474

RESULT 13

LOCUS CQ573675 2855 bp DNA linear PAT 02-FEB-2004

DEFINITION Sequence 1433 from Patent WO0171042.

ACCESSION CQ573675

VERSION CQ573675.1 GI:41637763

KEYWORDS

SOURCE Drosophila sp.

ORGANISM Drosophila sp.

REFERENCE 1

AUTHORS Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.

TITLE Detection kits, such as nucleic acid arrays, for detecting the expression of 10,000 or more Drosophila genes and uses thereof

JOURNAL Patent: WO 0171042-A 1433 27-SEP-2001;

FEATURES

source Location/Qualifiers

1..2855

/organism="Drosophila sp."

/mol_type="unassigned DNA"

/db_xref="taxon:7242"

ORIGIN

Query Match 30.3%; Score 733.4; DB 6; Length 2855;

Best Local Similarity 61.0%; Pred. No. 4.3e-155;

Matches 1302; Conservative 0; Mismatches 761; Indels 72; Gaps 4;

QY 121 CCTATACCTTCGCTCCATCTGCTAATCGCTGCTGGCGGATTTCTCTCGCCTCTT 190

Db 466 CTGTGTGAAGCTGGCCATTCATCTCTGGCAGCGTTTATCATTTGCCACAGCTGTG 525

QY 181 CGCGCTCATCGTTTCGAGTCGATTATCCATGAGTTTGATCCGTTGGTTCAACTACCGGC 240

Db 526 CTCTGTGTGCGATTGCAAGAGCGTAATCCATGAGTTGATCCGTTACTTCACTACCGCAC 585

QY 241 CACCGCCTACATGTTGCGAGATGTTGTGTACAACTTCTCACTGTTCCAGCAGCGCGC 300

Db 586 CACGCGTTCTTGGCGAGCAGGGCTTTTCAAGTTCCAACTTGGTTCCATGACCGCGC 645

QY 301 ATGGTATCCGCTCGGCAGGATTTGGCGGTACCGTCTATCCCGCGCTGATGATTACGTC 360

Db 646 CTGGTATCCCTTGGCGGCATCATCGCGGCACCATCTATCCCGGGCTGATGCTCACCTC 705

QY 361 CGCGGAATCCATTGGCTGCTGCAGTACTCAACATACCGGTCCATATTGTCGACATCTG 420

Db 706 GCGGCGCTGTACCGCCTGTATGTGCTCTCATGTGACCATCAACATACGGAACGTGTG 765

QY 421 CGTTTCTTGGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480

Db 766 CGTCTTCTTGGCGCGCTTCTCTCTCTGCTGACCACTGCTGACCTACGCGCTCACAA 825

QY 481 GGAGCTGTGGTCCGCGGCGCGCGCTCTTCCCGCCAGCTTATCCCATCTGCTGCTGG 540

Db 826 GGAGATACACAGCATGGAGCTGGAGCTGGTGGCGCGCTTTGATATCCATCTGTTCCGG 885

QY 541 CTACATCAGTAGTCCGTGGCTGGATCGTACGATAAAGAGGCGATTGCCATATTCGCGCT 600

Db 886 GTATATCTCTCGATCCGTGGCGGATCGTACGACATGAAGGACATCGCCATTTCTGCA 945

QY 601 GCAGTTCACCTACTCTCTGCTGGTGGCTCAGTGAAGCTGGATCCGCTGTTCTGCTGGC 660

Db 946 GCTCTTCCCTACTATTTGTGGATCAAGCGGTAAAGACGGGACGATCTTTTGGTGGC 1005

QY 661 CGCAGCGCTTTGCTCTACTTCTACATGGTGTCCGCTGGGCTGGCTACGTTGTTATCAT 720

Db 1006 TATGTCGCGCATTTGGCTTACTTATATGTTCTCTCGTGGGTGGTATGTTCTTCTGAT 1065

QY 721 CAACCTGATACCCCTGCAAGCTTCTGTTACTGCTCATTTATGGGAGGATATCGCGGCTCT 780

Db 1066 TAACTTAATCCGCTGCAAGCTGCTGGGCTGATGATCAACGGACGTTTCTCGACAGAT 1125

QY 781 GCTGACGAGCTACAGACCTTCTACATCTGGGACTGCTGTTCTCCATGAGATCCCTTT 840

Db 1126 CTACATAGCATACAGACGCTTACTGCTGGCGCACCACTCTGCTGATGAGATCTCGTT 1185

QY 841 CGTGGGATTCACACCGGATACGACAGTGAACACATGGCTGCGCTGGGAGTGTGTTGCT 900

Db 1186 TGTGGGATTCACACCGATCCAGAGCTCCGAACACATGCTGGCACTGGAACTTTGGCT 1245

QY 901 CTTATGCGCGTGGCGACCTTGGCCATTTGAGTGGCTGCTGCTGCGCAACAGATTCG 960

Db 1246 GTGCGAGATTCACGCTTTGCTGAGTATCTGCGCTCGGCACTTCCCAAGATCACTTGA 1305

QY 961 GAAGCTGTTATCGTGGCGGATTTGCTGGCGCTTGTGGCTTGTGTCGCTGCTGCTG 1020

Db 1306 TCTGCTCTTCAAGACGTTGTTTCCAGTGTGTTTGAAGTGTGTTGCTGCTGGGTACCT 1365

QY 1031 GCTCACAGCTGGCGCTTGTGGCGCGCTGAGTGAAGCTTCTACTCTGCTGGGATAC 1080

Db 1366 GCTCACGCTTACCGGGAAGTCTCCCGCTGGACCGGAGATTTTACTGCTACTAGATCC 1425

QY 1081 TGGTATGCCCAAGATCCACATTTCCATCATTTGCTGCGAGCATCGGCCACAC 1140

Db 1426 ATCTATGCCAAGATCACATTCATCCATCATCGCTCGGATCGAGCATCAGCTACATC 1485

QY 1141 TTGGTTCGTTCTTCTTGTGATCTGACATCTCTGTTGGCTTCCAGTGGGAGTGTG 1200

Db 1486 CTGTCGCTCTTCTTATTTTGTGATCTGAGATCTCTGCTCTTCTCTTCTCCGCTGGCTT 1545

QY 1201 TACTGTATCAAGCAGATCAACGACGAGCGCTTTCTGTTGCTGCTGAGCCATCAGTGC 1260

Db 1546 TTTCTGCTTTTCCAGCTGACGACTCGACATCTTATTTATTTGATGGCGTACCAG 1605

QY 1261 GGTTTATCTGCTGCTGTTGATGTTGTTGATGTTGACCCCTCACCGCGGTGTTGCTAT 1320

Db 1606 TATTTATTTCCCGGAGTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1665

QY 1321 GCTGCGCGAGTGGCTTTTCCGAGCTGTTGGATGTTGCTCTGCAAGAGATTCGCTAA 1380

Db 1666 GCTATCCGATTTGCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1719

QY 1381 GCGAATGGGCACAGCCATAAGCGCGCCACCGAAGTGGATGAAGCTGAGGATTCATTA 1440

Db 1720 -----TSCCGGAAGCTCATCGAAGCAGGA 1744

QY 1441 GAAGAAGAGCTGTACGCAAGGCTGGCAAGCTGTAAGCATGATGCCCCA 1500

Db 1745 GTAGAACG-----AAGCGCAGCACAAGAGCTGAACA 1779

QY 1501 CGAGATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560

Db 1780 GCAGACGGGAGCGCTGAAGAGCGAGTGGCAATTTGATTTGCTGCTGCTGCTGCTGCTGCT 1839

QY 1561 GTTATGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620

Db 1840 GCTTATAGTGTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1899

QY 1621 CATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680

Db 1900 TATTTGTTGAGTGC---CAGGTGCGACGATGGCGCGCATCATTTTCGATGACTTCCG 1956

QY 1681 AGAGCTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740

Db 1957 CGAGGCTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2016

QY 1741 GGATTCAGGATACAGATAGCGGAATGGCAACAGACGCTAGTGGATTAATATAC 1800

Db 2017 GGACTAGCGCTACCAGATAACAGCGCCATCGGCAATCGGACGATATTGGTGATACAATAC 2076
 QY 1801 GTGGAACAATAGTACATAGCGTGTGGCAGGCAATGCTCTCAACCGAGGAGATC 1860
 Db 2077 TTGGAACAACACATATATCGCGCTGGCCAGGCGATGGCTCTTCGGAGGAGAAAGC 2136
 QY 1861 CTACGAATATTGACATCTCTTACGCTGGACTACGTTTTGGTGATCTTTGGCGGTGTGAT 1920
 Db 2137 CTACGAGATAATAGGGAATCTGATGTGGACTACGTTCTCGTGATTTTCGGAGGCTCAC 2196
 QY 1921 CGGCTATCTCGGCATGATATCACAGTTCCTGTGGATGGTTCGGAATGCTGAGG--- 1977
 Db 2197 TGGTCTATCTCGGACATATCACAGTTCCTGTGGATGGTTCGGAATGCTGAGGACAC 2256
 QY 1978 AGAGCATCCCAAGACATTAAGGAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGT 2037
 Db 2257 GGATCGTGTGGCAGATCCGCAAAAGGACTACTATCGGCAACCGAGAGTTCGGAGT 2316
 QY 2038 AGATGCGGAGTCTCGGCGCTCGTCACTGCTTATGTAATTAAGCTACTACAG 2097
 Db 2317 GGACAGAGGGCTCACCACTGCTCAATTTGATGTACAGATGTGCTACTATCG 2376
 QY 2098 ATTCCGGAATTAAGTTGGAATACAGAGTCCATCTGGATATGATCGCACACGTAACGC 2157
 Db 2377 CTTTGGGCAATGTACAGGAAGTGGCAAGCCAGGCTAGCATCGAGTTCGTGGCGC 2436
 QY 2158 CGTCATTGGGAATAAGGACTTCGATCTGACCTACCTGAGGAGGCTACACACAGACA 2217
 Db 2437 CGAGATCGGCAACAGGACTTTGACATGGATGTCTGAGGAGGCGGTACACACGAGCA 2496
 QY 2218 CTGCTTGTGCTATGATGATGAGTGAAGAGCGCG 2252
 Db 2497 CTGCTGTGCGCATCTACAAGTTAAGGATCTGC 2531

RESULT 14

LOCUS BC046072 2608 bp mRNA linear VRT 30-JUN-2004
 DEFINITION Danio rerio integral membrane protein 1, mRNA (cdna clone MGC:56332 IMAGE:5603532), complete cds.
 ACCESSION BC046072
 VERSION 1
 KEYWORDS MGC
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio

REFERENCE 1 (bases 1 to 2608)
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Suetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, S., Moore, T., Max, S.L., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Casavant, T.L., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Ketchum, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shcherchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalski, U., Small, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 PUBMED 1247932
 REFERENCE 2 (bases 1 to 2608)
 AUTHORS Strausberg, R.

TITLE
JOURNAL

Direct Submission
 Submitted (31-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Sumio Sugano
 cDNA Library Preparation: Dr. Sumio Sugano
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

REMARK
COMMENT

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 103 Row: 6 Column: 9
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES

source

1..2608
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="MGC:56332 IMAGE:5603532"
 /tissue_type="Whole body, adult male"
 /clone_lib="Sugano SUD adult male"
 /lab_host="DH10B"
 /note="Vector: pME18S-FL3"

gene

1..2608
 /gene="itml"
 /note="Synonyms: cb437, wu:fb38c11, fb38c11"
 /db_xref="LocusID:266797"
 110..2227
 /gene="itml"
 /product="integral membrane protein 1"
 /protein_id="AAH46072.1"
 /db_xref="GI:28277990"

CDS

/translation="MTKLGLRLSYEKQDTLLKLLLSMAAVLSFSTLFSVLRFSSV
 IHEPDYFNRTTFLAEDEGFYKPHNFDRAWYPLGRIIGTITGLMITSAALYHV
 LHFPHITDIRNVCVFLAPLSPSTATVYFTKELKDAGALLAAMIAVVPGYISR
 SVAGSYDNEGIAFCMLITYMVIKAVKSGSIYVSSMCALAYFVWSMGVYFLNL
 IPLVILMLTGRFSHRIYVAVCTVCLGTILSMQISFVGPQVQSEHMAAFVFLGL
 COIHAFLVYLSKLNACQFVLPFSLVSLVGLILLVGLMLTGLTISLPTGTFYSL
 DPSYAKNIPILASVSEHPTNSSYFEDLQLLFVFPGLYFCFNNLSDATIFILY
 GTSYFSAVWVRLMLVLPVCMILSGISGVSVLTITFMKLDVSRDXTTKKQDSTY
 PKNSVAGSMILVWAFLLITYFTSTVTSVSEVSSPILSVARGGDSRIIPDFREA
 YVLRHNPEDAKVMSWMDYGYQTAMANETILVDNNTWNTHSIRVQAMASTEKA
 YELRDELVSFLVIFGLTGYSSDDINKFLWVRIGISTDGRHKEHDYVTPGEP
 VDRGSPVLLNCLWYKMYRFQGVYTEAKRPGYDRVRNBSIGNKDFELDLVBAY
 TTEHLVRIYKVKLDLNRGLSRT"

ORIGIN

Query Match 29.8%; Score 720.2; DB 5; Length 2608;
 Best Local Similarity 60.8%; Pred. No. 4.1e-152;
 Matches 1299; Conservative 0; Mismatches 743; Indels 93; Gaps 4;
 QY 118 CAGCTTAATCACTTCGCGATCCCTGCTAATCGCTGGCGCGGATTTTCCTCTCGCCT 177
 Db 154 CACCTGTGTGAGTGTCTCATCTGTCTATGCGACCGTCTCTCTCCACGACT 213
 QY 178 CTTGCGCGTATCCGTTTCGAGTCGATTCATCGATTCGTCGTTGTTCACTACCG 237
 Db 214 GTTTTCTGTGCTTCGGTTCGAGAGCGTATCCACGAGTTTGTATTCGTAATACCG 273
 QY 238 GGCCACCGCGCTACATGTTGTCAGAAATGGTTGGTACAACTTCCTCAACTGGTTCCGACGAGCG 297

TITLE

JOURNAL
 PUBMED
 REFERENCE 2 (bases 1 to 2608)
 AUTHORS

Db 274 GACCACTCGCTCTCTGGCAGAGAGGAGATTTTACAAATTCATACCTGCTGATGATCG 333
QY 298 CGCATGGTATCGGCTCGGACGATTTGGCGGTACCGTCTATCCCGCGCTGATGATTAC 357
Db 334 GCGTTGGTATCCACTTGGGAGGATTTATGGAGGACCATCTATCTCGGGTTGATGATCAC 393
QY 358 GTCCGGCGGAATCCATTTGGCTGCTGCAAGTACTCAACATACCGGTCCATATTCGTGACAT 417
Db 394 ATCAGACGCTCTATATCAGTGTGCACTTCTTCACATCACCATTGACATCCGTAATGT 453
QY 418 CTGCGTGTCTCGGCGCGATCTTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGAC 477
Db 454 CTGCTGTCTCGCT 513
QY 478 CAAAGAGCTGTGGTCCGGCGCGCGGCTCTCTCGCGCGGCTCTCTCTCTCTCTCTCTCT 537
Db 514 CAAAGAGCTGAAAGATCGGCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 573
QY 538 TCGCTACATCAGTAGGTCTGGTGGTGGATCTGATACGATACGAGGGCATTCCTCATATTCGC 597
Db 574 TGGATACATCTCCCGTCTCTGTCGAGGCTCTTATGATAATGAAGGAATTCGCAATTTCTG 633
QY 598 CCGCAGTTCACCTACTTCTCTGCGGTGCGCTCAGTGAAGACTGATCGGTCTCTGCTC 657
Db 634 CATGCTGTGAGTACTACATGTGATCAAGGAGTGAAGTGGGCTCTCAATTTATTTGGTC 693
QY 658 GCGCGCAGCGGCTTTGCTCTACTTCTACATGCTGCTCGGCTGGGCTGGCTGCTGCTCAT 717
Db 694 ATCCATGTGTGACCTGGCATATTTTACATGCTGCTCTCTCTCTCTCTCTCTCTCTCT 753
QY 718 CATCAACCTGATACCCCTGACGCTCTGCTACTCTCATTTATGGGAGTACTCGCGCG 777
Db 754 GATAAATCTCATCTCTGACGCTGCTGCTGATGCTCAGCGCGGCTCTCTCTCATCG 813
QY 778 TCTGCTGACCACTACAGCACTTCTACATCTGGGACTGCTGTTCTCCATGACAGATCC 837
Db 814 CATCTATGGGCTACTGACAGTCTACTGCTGGGACCATCTCTCCATGACAGATCTC 873
QY 838 CTTCGTGGATTCACACCGATACGACAGTGAACATGCTGCTGGGAGTACTCGCGCG 897
Db 874 TTTTGTGATTCACAGCCGCTCCAGTCTCTGAAACATGCTGCTTTTGGAGTGTGG 933
QY 898 GCTCTTATGGCGGTGGCCACTTTGGCCATTTTCAGTCCGCTGCTGCTGCGCAACAGATT 957
Db 934 TCTGTGCCAGATTCACGCTTTTGTGATTTACTCTGGGAGCAAACTGAACCGCAGCAGTT 993
QY 958 CCGGAGCTGTTCATCTGCGGATTTCTGCTGGGCTGGGCTGCTTTGCTGCGCGCT 1017
Db 994 TGAGGTGTGTTTCAGAGTGTCTATTTCACTGCTGGGATCATCTGCTGCTGCTGGCGG 1053
QY 1018 GGTGCTCACCATGCTGGCGGTGTGGCCCGCTGGAGTGGAGCTTCTACTCTGCTGGGA 1077
Db 1054 GTTGTCTATGCTGACAGGAAATCTCCCTTTGACTGGCGTTTCTACTCTCTGCTGGA 1113
QY 1078 TACTGGCTACGCAAGATCCATTCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1137
Db 1114 TCCATGTCAGGCAAGAACATTCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1173
QY 1138 CACTTGGTCTGCTTCTCTTCTGATCTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
Db 1174 CACATGGTCTCTCTACTTCTGACCTCCAGCTTCTGCTCTCATGTTTCCAGTTGGCT 1233
QY 1198 GTGGTACTGCATACAGAGATCAACGACGCGGCTTTTGGTGTGCTGCTGCTGCTGCTGCT 1257
Db 1234 TTAATGCTGCTTCAACACCTCTCTGATGCGCACGATTTTATCATCATGATGCTGCTGCT 1293
QY 1258 TCGGCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
Db 1294 CAGCATGATCTTCTCAGCTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1353
QY 1318 CATGCTGCGGAGTGGGCTTTTTCGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377

Db 1354 CATCTGCTGCGCATTTGGGCTCTC----- 1377
QY 1378 TAAACGAATGGGACACGCCATAAGCCAGCCACCGAAGTGGATGAAGCTGAGGATTCAT 1437
Db 1378 ---TCAGTCTCTACCACTTTCATGAAGAACCTGGATGTCAAGTCCAGCCGACCAAGAAAC 1434
QY 1438 TGAGAAGAAGCGGTGTACGACAAGGCTGGCAAGCTGAAGCATCTTAAGCATGATGCTG 1497
Db 1435 CAAGAGCAGCAGAGCTCCACCTACCCATCAAG----- 1468
QY 1498 CCAGCAGGATACTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1557
Db 1469 -----AACGAAGTGGCAAGTGGGATGATTTCTAGTGTGCTGCT 1506
QY 1558 GCTGTTGATGATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617
Db 1507 CTTCTCTATCATACATGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1566
QY 1618 CTCGATGCTGCTGCTGCTTTCACCAACAGTCAAGATGGATCCCGCAACATTTTAGACGATT 1677
Db 1567 CTCGATGCTGCTTCTGCTC-----GCGGTGGCGAGCGGCGGATCATCTTTGATGACTT 1623
QY 1678 CAGAGGCTTACTGCTGCTTTCGCAAGACACTGCGGATGATGCTGCTGCTGCTGCTGCT 1737
Db 1624 CAGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1683
QY 1738 GTGGATTCAGGATACAGATAGCGGATGGCAACAGCAACAGCACTGCTGCTGCTGCTGCT 1797
Db 1684 GTGGATTCAGGATACAGATAGCGGATGGCAACAGCAACAGCACTGCTGCTGCTGCTGCT 1743
QY 1798 TACGTGGAACAAATAGTCACATAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1857
Db 1744 CACTTGGATTAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1803
QY 1858 GTCTACGAATTAAGATCTCTTGAAGTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1917
Db 1804 ASCCTATGAGATTAAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1863
QY 1918 GATCGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1977
Db 1864 GACCGATATTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1923
QY 1978 ---AGAGCATCCCAAGACATTAAGGAAGCACTTACCGACCGGCTGCTGCTGCTGCT 2034
Db 1924 CACAGACACAGGAGGACATTAAGGAGCACTTATTAACGCCCACTGCTGCTGCTGCTGCT 1983
QY 2035 GGTAGATCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2094
Db 1984 TGTGACCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2043
QY 2095 CAGATTCCGGGAATTGAAGTGGATACAGAGTCCATCTGGATATGATGCTGCTGCTGCTGCT 2154
Db 2044 TCGCTTCCGGCAGGCTTACACCGAAGCGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 2103
QY 2155 CGCGCTCATTTGGGATTAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2214
Db 2104 TGCTGAGATTGGCAATTAAGGCTTGAAGTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2163
QY 2215 AACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2249
Db 2164 GCACTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2198

RESULT 15

BC063234

LOCUS

DEFINITION

IMAGE:7001298), complete cds.

ACCESSION

BC063234.1

VERSION

GI:38649116

KEYWORDS

MGC

SOURCE

Danio rerio (zebrafish)

ORGANISM

Danio rerio

BC063234 2567 bp mRNA linear VRT 30-JUN-2004

Danio rerio integral membrane protein 1, mRNA (cDNA clone MGC:77796

IMAGE:7001298), complete cds.

ACCESSION BC063234

VERSION BC063234.1

GI:38649116

KEYWORDS MGC

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 2567)
 Srausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shemen, C.M., Schuler, G.D., Altschul, S.F., Zebner, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Partner, A.A., Rubin, G.W., Hong, L., Scaletton, M., Soares, W.B., Ronaldo, M.P., Casavant, T.L., Schaefer, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Murray, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.O., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL
 PUBLISHED
 12477932

2 (bases 1 to 2567)
 Srausberg, R.
 Direct Submission
 Submitted (02-DEC-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-rc@mail.nih.gov
 Tissue Procurement: Len Zon, Harvard
 cDNA Library Preparation: Open Biosystems
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Allison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu, Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smalish, Jeff Stott, Miranda Tsai, George Yang, Jacquie Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 147 Row: 1 Column: 3
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis. Similarity but not identity to protein.

FEATURES

Location/Qualifiers
 1..2567
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="MGC:77796 IMAGE:7001298"
 /tissue_type="Whole, adult zebrafish"
 /clone_lib="NIH_ZGC_10"
 /lab_hosts="DH10B"
 /note="vector: pexprs1"
 1..2567
 /gene="itnl1"
 /note="synonyms: cb437, wu:fb38c11, fb38c11"
 /db_xref="LocusID:266797"
 70..2187
 /gene="itml"

gene

CDS

/codon_start=1
 /product="Integral membrane protein 1"
 /protein_id="AAH63234.1"
 /db_xref="GI:38649117"
 /db_xref="LocusID:266797"
 /translation="MTKGLFLRLSYEKQDTLLKLLILSMAVLSTFRLPSVLRPSV
 IHEFPYNYVTRFLAEDEGFKPHNWFDDRAWYPLGRIGGTIYGLMITSAALYHV
 LHFPHITIDIRNVCVELAPLPSFTAIYVFTKELADAGALLAAMTAVVPGYLSR
 SVAGSYDNREGIAJECMLITYMYMIKAVKSGIYKSSMCALAYFYWSSRGVYFLNL
 IFLHVLVLMJGRFSRIYVAICTYVGLGILLSNQISFVGFQVQSEHWAQFVGL
 COIHAFVILRSKLNAAQFVLRPSVLSLVGILLSSVGLMLTKISWTRGFRFSL
 DPSYAKNIPITIASVSEHQPTWSSYFDLQVLFMPVGLYTYCPNNLSDAITFIIMY
 GVTSYFSAVNRMLVLAQVMSLIGISQVLTPEKNDLVSRDKPKTKQKQDSTY
 PIKNEVASCMILVMAFPLITYTHSTWTSYSEAYSSPSVLSARGGSGSRIIFDREA
 YWLRNTPEDAKWMSWDYGYQTAMANRTILVDNNTWNTHTSRVQAMATBEKA
 VEIRBELDSYVLVIFGLTGYSDDDINKELMWVRIGSDTDGTHIKEDHYTPTGEF
 RVDEGSLVLLNCLMYKMCYRQGVTEAKRPPGYDRVNAIGNKDFELDVLEBAY
 TTEHVLVRIYKVKOLDNRGLSRT"

ORIGIN

Query Match 29.7%; Score 718.6; DB 5; Length 2567;
 Best Local Similarity 60.8%; Pred. No. 9.5e-152;
 Matches 1298; Conservative 0; Mismatches 744; Indels 93; Gaps 4;

QY	118	CAGCCTATACCTTCGGCATCTGCTAATCGCTGCGCTGCGCGGATTTTCCTCGCCT	177
DB	114	CACCTGCTGAAGCTGCTCATCTCTGTCTATGGAGCGCTCTCTCTCTCCACAGACT	173
QY	178	CTTCGCGCTCATCCGTTTCGAGTCGATATCATGAGTTTTCGCGGTGTTCAACCG	237
DB	174	GTTTTCTGTGCTCGTTTCGAGAGCGCTCATCCAGAGTTTGATCCGTTATTTCAATACCG	233
QY	238	GGCCACCGCTACATGCTGCAGAAATGTTGTGTACAACTTCTCACTGTTTCACAGCG	297
DB	234	GACCACTGCTTCTCGGAGAGAGGAGATTTTACAAATTCATATCTGTTGATGATCG	293
QY	298	CGCATGGTATCCGCTCGGCAGAGATTGTGGCGGTACCGTCTATCCCGCGCTGATGATAC	357
DB	294	GGCTTGGTATCCACTTGGAGGATTATTGGAGGACCATCTATCTCTGGTGTGATGATCAC	353
QY	358	GTCCGGCGGAATCCATTGGCTGCTGCAGTACTCAACATACCGGTCCATTTCTGCACAT	417
DB	354	ATCAGAGCTTATATATACAGTGTGCTACTTTCCATCATCACAATGACATCGTATGTT	413
QY	418	CTGCGTGTTCCTGGCGCGGATCTTCAGTGGCGTACCTCCATCTCCACCTACCTGCTGAC	477
DB	414	CTGTGTGTTCTCGCT	473
QY	478	CAAGGAGCTGTGTCCTCGCGCGCGCGCTCTTCGCGCGGAGCTTCATCGCCATCGGCC	537
DB	474	CAAGAGCTGAAGATGAGCGGCTGGGCTTCTGGCTGCGCTATGATGATGATGATGATGATG	533
QY	538	TGGTACATCATAGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	597
DB	534	TGATACATCTCCGTTCTGTGGCAGGCTCTTATGATAATGAAGAAATTCGCAATTTCTG	593
QY	598	CTGCGAGTTCACCTACTTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	657
DB	594	CATGCTGTGAGTACTACATGATGATCAAGGAGTAGTGAATCGGCTCTCAATTTATGCTC	653
QY	658	GGCGGAGCGGCTTGTCTCTACTTTCTACATGGTGTCCGCTGGGCTGGCTACGTTTCAT	717
DB	654	ATCCATGTGTGATTTGGCATTTATTTTACATGGTGTCTCTCTCTCTCTCTCTCTCTCTCT	713
QY	718	CATCAACCTGATACCCCTGCGAGCTCTTCTGCTACTCTCTCATTTATGGGAGGTAATCGCGCG	777
DB	714	GATTAACCTCATCTCTCTCTGACGCTGCTGTTGATGCTTACCGCGCGGTTCTCTCATCG	773
QY	778	TCCTGCTGACGACTACAGCACCTTTCTACATCTCTGGGACTGCTGTTCTCCATGAGATGCC	837
DB	774	CATCTATGTGGGCTACTGACAGTCTACTGCTGGGACCATCTCTCTCTCCATGAGATCTC	833
QY	838	CTTCGTGGGATTCACACCGATACGACACCATGACATGCTCTGCTGCGGTGGGAGTGTGT	897

Blank
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 20:50:43 ; Search time 1108.82 Seconds
(without alignments)
11442.663 Million cell updates/sec

Title: US-10-028-384-7

Perfect score: 2417

Sequence: 1 tctaaagcgaagaagtgtctg.....ccaaaaaataaaaaaa 2417

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N Geneseq 23Sep04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002as:*

7: geneseqn2002bs:*

8: geneseqn2003as:*

9: geneseqn2003bs:*

10: geneseqn2003cs:*

11: geneseqn2003ds:*

12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2417	100.0	2417	10	ADD94789	ADD94789 Drosophila
2	2410.4	99.7	2699	4	ABL13247	ABL13247 Drosophila
3	2147.4	88.8	4922	4	ABL13246	ABL13246 Drosophila
4	960.6	39.7	2710	10	ADD94785	ADD94785 Mouse SIM
5	955.8	39.5	2481	10	ADD94783	ADD94783 Human SIM
6	733.4	30.3	2855	4	ABL02795	ABL02795 Drosophila
7	696	28.8	2953	4	ABL18224	ABL18224 Drosophila
8	635.4	26.3	2232	8	ABT20789	ABT20789 Aspergill
9	632.8	26.2	2472	10	ADD94793	ADD94793 Human ITM
10	632.8	26.2	2472	10	ADH28844	ADH28844 Human chr
11	631.2	26.1	3094	10	ADD94791	ADD94791 Mouse chr
12	620.2	25.7	2466	10	ADBE6900	ADBE6900 C. neofor
13	615.6	25.5	2760	5	ABV24502	ABV24502 Human pro
14	562	23.3	1848	8	ABT18969	ABT18969 Aspergill
15	521.8	21.6	2733	10	ADD94787	ADD94787 Yeast STT
16	521.6	21.6	2000	10	ACC61293	ACC61293 Gene sequ
17	521.6	21.6	2000	10	ADK63659	ADK63659 Disease t
18	519.8	21.5	2256	6	ABZ32036	ABZ32036 Candida a
19	501.8	20.8	1664	4	AAK94164	AAK94164 Human ful
20	501.8	20.8	1864	12	ADL30661	ADL30661 Full leng
21	497	20.6	1969	8	ABT18375	ABT18375 Aspergill

22	497	20.6	2603	8	ABT20191	ABT20191 Aspergill
23	497	20.6	3969	8	ABT17781	ABT17781 Aspergill
24	497	20.6	4603	8	ABT19595	ABT19595 Aspergill
25	467	19.3	1209	4	AAH33264	AAH33264 Human col
26	467	19.3	1209	6	ABL89850	ABL89850 Human col
27	466.4	19.3	1543	2	AAH85055	AAH85055 Human sec
28	466.4	19.3	1543	8	ACD18981	ACD18981 Novel hum
29	466.4	19.3	1543	12	ADG78372	ADG78372 Human sec
30	466.4	19.3	1543	12	ADN60663	ADN60663 Human sec
31	444.4	18.4	1371	12	ADP28508	ADP28508 Human sec
32	414.6	17.2	2510	5	RAF93772	RAF93772 Human CDN
33	414.6	17.2	2546	2	RAV44866	RAV44866 Clone CTS
34	414.6	17.2	2546	5	AAF98463	AAF98463 Human CDN
35	408.4	16.9	2785	4	ABL18208	ABL18208 Drosophila
36	403	16.7	2537	4	ADN08289	ADN08289 Human sec
37	403	16.7	2547	4	ADN08315	ADN08315 Human sec
38	403	16.7	2660	6	ABQ54750	ABQ54750 Human ova
39	392.4	16.2	2284	4	AAH18021	AAH18021 Human CDN
40	373.8	15.5	2882	10	ADB69539	ADB69539 C. neofor
41	373.8	15.5	4738	10	ADB69178	ADB69178 C. neofor
42	356.2	14.7	787	4	AAH07526	AAH07526 Human CDN
43	321.2	13.3	764	5	AAF92968	AAF92968 Primer sp
44	310.8	12.9	1114	4	AAH95794	AAH95794 Human pro
45	279.6	11.6	507	3	AAH01967	AAH01967 Human sec

ALIGNMENTS

RESULT 1

ADD94789

ID ADD94789 standard; DNA; 2417 BP.

AC ADD94789;

DT 29-JAN-2004 (first entry)

DE Drosophila melanogaster SIT3 gene sequence.

KW source of immunodominant MHC-associated peptide; SIMP; MHC;

KW major histocompatibility complex; human leukocyte antigen; HLA;

KW cytotoxic; immunosuppressive; antitense therapy; gene therapy; cancer;

KW lung cancer; intestine cancer; sarcoma; prostate cancer;

KW testicular cancer; breast cancer; melanoma; pancreatic cancer;

KW haematological cancer; immune response; lymphoid cell proliferation;

KW autoimmune disease; transplant rejection; SIMP-derived peptide;

KW fruit fly; gene; ds; SIT3.

OS Drosophila melanogaster.

PN WO2003054008-A2.

XX 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

XX 20-DEC-2001; 2001US-00028384.

XX (COMP-) COMPATIGENE INC.

XX Perreault C, McBride K;

XX WPI: 2003-559122/52.

XX P-FSDB; ADD94790.

XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancer, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.

XX Claim 6; SEQ ID NO 7; 66pp; English.

XX This invention relates to a novel isolated or purified human protein,

CC

Query Match	100.0%	Score	2417;	DB	10;	Length	2417;
Best Local Similarity	100.0%	Pred.	No. 0;				
Matches	2417;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	TCTAAGCCGAAGAATGTCGTGCTGCAATTCAGATCGGTATATAATTTTCAGATTACTGGCTG	60				
DB	1	TCTAAGCCGAAGAATGTCGTGCTGCAATTCAGATCGGTATATAATTTTCAGATTACTGGCTG	60				
QY	61	GAATTTGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGAG	120				
DB	61	GAATTTGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGAG	120				
QY	121	CCTAATCACCTTCGCCATCCTGCTAAATCGGCTGGCTGGCGGATTTTCCTTCGGCTCTT	180				
DB	121	CCTAATCACCTTCGCCATCCTGCTAAATCGGCTGGCTGGCGGATTTTCCTTCGGCTCTT	180				
QY	181	CGCGTCAATCCGTTTCAGATCGAATATCAATGAATTTGATTCGTGGTTCAACTACCGGCG	240				
DB	181	CGCGTCAATCCGTTTCAGATCGAATATCAATGAATTTGATTCGTGGTTCAACTACCGGCG	240				
QY	241	CACCGCTTACATGGTCAGAAATGGTGGTGAACAATTCCTCAACTGGTTCGACGACGCGCG	300				
DB	241	CACCGCTTACATGGTGGCAGAAATGGTGGTGAACAATTCCTCAACTGGTTCGACGACGCGCG	300				
QY	301	ATGGTATCCGCTCGGACAGAAATGGGCGGTACCGTCTATCCGGGCTGATGATTACGTC	360				
DB	301	ATGGTATCCGCTCGGACAGAAATGGGCGGTACCGTCTATCCGGGCTGATGATTACGTC	360				
QY	361	CGGCGGAATCCATTGGGCTGCTGCACGTACTCAACATACCGGTCCATATTTCGTGACATCTG	420				
DB	361	CGGCGGAATCCATTGGGCTGCTGCACGTACTCAACATACCGGTCCATATTTCGTGACATCTG	420				
QY	421	GTGTTCTCGGCGCGAATCTTCAGTGGCTGAACCTCATCTCCACTACTGCTGACCAA	480				
DB	421	GTGTTCTCGGCGCGAATCTTCAGTGGCTGAACCTCATCTCCACTACTGCTGACCAA	480				
QY	481	GGAGCTGTGGTTCGCGGCGCGCGGCTCTTTCCGCGCGCAGCTTCATCGGCATCGTGCCTGG	540				
DB	481	GGAGCTGTGGTTCGCGGCGCGCGGCTCTTTCCGCGCGCAGCTTCATCGGCATCGTGCCTGG	540				
QY	541	CTACATCAGTAGTTCGTGGCTGGATCGTACGATTAACGAGGGCATTCGATTCGCCCT	600				
DB	541	CTACATCAGTAGTTCGTGGCTGGATCGTACGATTAACGAGGGCATTCGATTCGCCCT	600				
QY	601	GCAGTTCACTTACTTCCTGTGGTGGCGCTCAGTGAAGACTTGGATCCGTGTTCTGCTGGCG	660				
DB	601	GCAGTTCACTTACTTCCTGTGGTGGCGCTCAGTGAAGACTTGGATCCGTGTTCTGCTGGCG	660				
QY	661	CGCAGCGGCTTTGTCTTACTTCTACATGGTGTCCGCTGGGTGGCTACGTTTTCATCAT	720				
DB	661	CGCAGCGGCTTTGTCTTACTTCTACATGGTGTCCGCTGGGTGGCTACGTTTTCATCAT	720				
QY	721	CAACCTGATACCCCTGGACGCTTTTCGTACTGCTCATTTATGGGCAGGTACTCCGCCGCTCT	780				

Db 730 CAACCTGATAGCCCTTCACGCTCTTCGTAAGTCTCAATATGGGAGGTAATCGCCGCTCT 789
Qy 781 GCTGACAGCTACAGACCTTCTACATCCCTGGGACTGCTGTTCTCCATCGACATCCCTT 840
Db 790 GCTGACAGCTACAGACCTTCTACATCCCTGGGACTGCTGTTCTCCATCGACATCCCTT 849
Qy 841 GCTGGGATTCACACCGATACGCAACAGTGAACATGCTGCTGCTGGGAGTGTGCTGCT 900
Db 850 GCTGGGATTCACACCGATACGCAACAGTGAACATGCTGCTGCTGGGAGTGTGCTGCT 909
Qy 901 CTTATAGCCCTGGCCACTTGGCCACTTGGCCACTTGGCCACTTGGCCACTTGGCCACT 960
Db 910 CTTATAGCCCTGGCCACTTGGCCACTTGGCCACTTGGCCACTTGGCCACTTGGCCACT 969
Qy 961 GAAGCTGTTCAATGCTGGGAGTGTGCTGCTGGGAGTGTGCTGCTGGGAGTGTGCTGCT 1020
Db 970 GAAGCTGTTCAATGCTGGGAGTGTGCTGCTGGGAGTGTGCTGCTGGGAGTGTGCTGCT 1029
Qy 1021 GCTCAACATGCTGGGAGTGTGCTGCTGGGAGTGTGCTGCTGGGAGTGTGCTGCTGGG 1080
Db 1030 GCTCAACATGCTGGGAGTGTGCTGCTGGGAGTGTGCTGCTGGGAGTGTGCTGCTGGG 1089
Qy 1081 TGGCTACGCAAGATTCACATCCCATTCATTCATTCATTCATTCATTCATTCATTCAT 1140
Db 1090 TGGCTACGCAAGATTCACATCCCATTCATTCATTCATTCATTCATTCATTCATTCAT 1149
Qy 1141 TTGGTCTCTGCT 1200
Db 1150 TTGGTCTCTGCT 1209
Qy 1201 GTACTGATCAAGAGATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1210 GTACTGATCAAGAGATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1269
Qy 1261 GCTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 1270 GCTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1329
Qy 1321 GCTGCTGGGAGTGGCTTTTGGGAGTGGTGGAGTGGTGGAGTGGTGGAGTGGTGGAG 1380
Db 1330 GCTGCTGGGAGTGGCTTTTGGGAGTGGTGGAGTGGTGGAGTGGTGGAGTGGTGGAG 1389
Qy 1381 GCGAATGGGACAGCATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
Db 1390 GCGAATGGGACAGCATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1449
Qy 1441 GAAGAAGAGCTGTACAGAGAGGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAG 1500
Db 1450 GAAGAAGAGCTGTACAGAGAGGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAG 1509
Qy 1501 GCAGATGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Db 1510 GCAGATGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1569
Qy 1561 GTTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Db 1570 GTTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1629
Qy 1621 CATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Db 1630 CATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1689
Qy 1681 AGAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 1690 AGAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1749
Qy 1741 GGATTACGGATACAGATAGCGGAGTGGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Db 1750 GGATTACGGATACAGATAGCGGAGTGGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1809
Qy 1801 GTGGACATAGTACATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860

Db 1810 GTGGACATAGTACATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1869
Qy 1861 CTACGAAATTTATGACATCTCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
Db 1870 CTACGAAATTTATGACATCTCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1929
Qy 1921 CGCTATTCTGGGATGATATCAACAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
Db 1930 CGCTATTCTGGGATGATATCAACAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1989
Qy 1981 GCATCCCAAGGACATTAAGGAAAGCGATTTACTTTACCGACCGGCTGAAATTCAG 2040
Db 1990 GCATCCCAAGGACATTAAGGAAAGCGATTTACTTTACCGACCGGCTGAAATTCAG 2049
Qy 2041 TGCAGAGTGTCTCGGCCCTGCTCAACTGCTTATGTAGAAATTAAGTACTACAGATT 2100
Db 2050 TGCAGAGTGTCTCGGCCCTGCTCAACTGCTTATGTAGAAATTAAGTACTACAGATT 2109
Qy 2101 CGGGGAATTCAGGATTTGGACTACAGAGTCCATCTGGATATGATCGCACAGTAAAC 2160
Db 2110 CGGGGAATTCAGGATTTGGACTACAGAGTCCATCTGGATATGATCGCACAGTAAAC 2169
Qy 2161 CATTTGCTGATAGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
Db 2170 CATTTGCTGATAGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2229
Qy 2221 GCTTGTCTGATCTATAGGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
Db 2230 GCTTGTCTGATCTATAGGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2289
Qy 2281 CAAGGAGAGAGATTCCTTCCAGCAAACTTCATTTCCGAGAGAGAGTCTGAGCGTCC 2340
Db 2290 CAAGGAGAGAGATTCCTTCCAGCAAACTTCATTTCCGAGAGAGAGTCTGAGCGTCC 2349
Qy 2341 GGCTACATACAGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
Db 2350 GGCTACATACAGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2409
Qy 2401 AAAAAAAAAA 2412
Db 2410 AAAAAAAAAA 2421

RESULT 3
ABL13246/s
ID ABL13246 standard; cDNA; 4922 BP.
XX ABL13246;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX OS
XX PN
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-FSDS; ABB69143.
XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1: SEQ ID NO 34220; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
CC AB572072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIFO. int/pub/published_pct_sequences
XX
SQ Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;
Query Match 88.8%; Score 2147.4; DB 4; Length 4922;
Best Local Similarity 91.5%; Pred. No. 0;
Matches 2411; Conservative 0; Mismatches 1; Indels 223; Gaps 4;
QY 1 TCTAAGCGAAGATGTGCTGCTATTCAGATCGTTTATAATTTTCGAGTTACTGCTG 60
DB 3913 TCTAAGCGAAGATGTGCTGCTATTCAGATCGTTTATAATTTTCGAGTTACTGCTG 3854
QY 61 GAATTGGGACATGAATCGGACGCCGAGATGCTGAACAGCAAGTGTGCTGCTACAGCAG 120
DB 3853 GAATTGGGACATGAATCGGACGCCGAGATGCTGAACAGCAAGTGTGCTGCTACAGCAG 3794
QY 121 CTTATACCTTCGCTCATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTGCGCTCTT 180
DB 3793 CTTATACCTTCGCTCATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTGCGCTCTT 3734
QY 181 CGCGTTCATCGTTTCGAGTCGATTAATCGATGAGTTGATCCGTTGTTCAATACCGGGC 240
DB 3733 CGCGTTCATCGTTTCGAGTCGATTAATCGATGAGTTGATCCGTTGTTCAATACCGGGC 3674
QY 241 CACCGCTACATGTTGCGAAGTGTGTGACACTTCCTCACTGTTTCGACGAGCGCGC 300
DB 3673 CACCGCTACATGTTGCGAAGTGTGTGACACTTCCTCACTGTTTCGACGAGCGCGC 3614
QY 301 ATGGTATCGCTCGGCGAGATTTGGCGGTACCGTCTATCCGCGCTGATGATACGTC 360
DB 3613 ATGGTATCGCTCGGCGAGATTTGGCGGTACCGTCTATCCGCGCTGATGATACGTC 3554
QY 361 CGCGGAATCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 3553 CGCGGAATCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3494
QY 421 CGTGTTCCTGCGCGGATTTGAGTGGCTGACCTTCATCTCCACCTACCTGCTGACCAA 480
DB 3493 CGTGTTCCTGCGCGGATTTGAGTGGCTGACCTTCATCTCCACCTACCTGCTGACCAA 3434
QY 481 GAGCTGTGTGCTGCGCGGCGCGGCTCTTGGCGCGGAGTTTCATCGCATCTGCGCTGG 540
DB 3433 GAGCTGTGTGCTGCGCGGCGCGGCTCTTGGCGCGGAGTTTCATCGCATCTGCGCTGG 3374
QY 541 CTATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
DB 3373 CTATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3314
QY 601 GCAGTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
DB 3313 GCAGTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3254
QY 661 CGCAGCGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
DB 3253 CGCAGCGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3194
QY 721 CAACCTGATACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780

DB 3193 CAACCTGATACCCCTGACAGCTTTCGATCTGCTCATTTATGGGAGGATCTCGCGGCTCT 3134
QY 781 GCTGACCAAGCTACAGCAACCTTCTACATCTCTGGAGCTGCTTTTCCATGAGATCCCTTT 840
DB 3133 GCTGACCAAGCTACAGCAACCTTCTACATCTCTGGAGCTGCTTTTCCATGAGATCCCTTT 3074
QY 841 CGTGGGATTTCAACCGGATAGCACCAGTGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 3073 CGTGGGATTTCAACCGGATAGCACCAGTGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 3014
QY 901 CTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
DB 3013 CTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2954
QY 961 GAAGCTGTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
DB 2953 GAAGCTGTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2894
QY 1021 GCTCACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
DB 2893 GCTCACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2834
QY 1081 TGGCTACGCCCAAGATCCACATTCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 2833 TGGCTACGCCCAAGATCCACATTCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2774
QY 1141 TTGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
DB 2773 TTGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2714
QY 1201 GTACTGCTATCAAGCAGATCAACGACGAGCGCTTTTCTGCTGCTGCTGCTGCTGCTGCTG 1260
DB 2713 GTACTGCTATCAAGCAGATCAACGACGAGCGCTTTTCTGCTGCTGCTGCTGCTGCTGCTG 2654
QY 1261 GGTTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
DB 2653 GGTTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2594
QY 1321 GCTGCGCGGAGTGGCTTTTTCGAGCTGTTGGATGTTTCTTCAAGAGGATTCGCTTAA 1380
DB 2593 GCTGCGCGGAGTGGCTTTTTCGAGCTGTTGGATGTTTCTTCAAGAGGATTCGCTTAA 2534
QY 1381 CGGAATGGGACAGCAGATTAAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
DB 2533 CGGAATGGGACAGCAGATTAAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2474
QY 1441 GAAAGAGAGCTGTACGACA----- 1460
DB 2473 GAAAGAGAGCTGTACGACAAGGTGAGTCTTACTAACACACATCCATGGGATTTGATTT 2414
QY 1461 -----AGGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG 1501
DB 2413 TAAATATACATCGCATTTTTCAGGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG 2354
QY 1502 CAGGATACCTGCGCTCAGCTCCCACTCAAGAGTATTTTATTTTGGCCGTTCTTAATGCTG 1561
DB 2353 CAGGATACCTGCGCTCAGCTCCCACTCAAGAGTATTTTATTTTGGCCGTTCTTAATGCTG 2294
QY 1562 TTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1621
DB 2293 TTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2234
QY 1622 ATTGCTTGGCTTTTCCCAACAGTCAAGATG----- 1652
DB 2233 ATTGCTTGGCTTTTCCCAACAGTCAAGATGTTGGTAAAGCAACATAGCATCATGCTTCT 2174
QY 1653 -----GATCCCGCAAGATTTTATAGAGATTTTATAGAGAG 1684
DB 2173 TTAACCTATCTAACCTTTTCTCCACTAGATCCCGCAACATTTTATAGAGATTTTATAGAG 2114
QY 1685 GCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1744

Db 2113 GCTTACTACTGGCTTTCGCGAACAACCTGCGGATGATGCTCGCGTTAGTCTTGTGGGAT 2054
 Qy 1745 TACGGATACAGATAGCGGGAATGGCAAAACAGAACGACGCTAGTGGATTAATAATACGTGG 1804
 Db 2053 TAGCGATACAGATAGCGGGAATGGCAAAACAGAACGACGCTAGTGGATTAATAATACGTGG 1994
 Qy 1805 AACATAGTACATAGCGTGGTGGGAGGCAATGCTTCAACCGAGGAGAGTCTTAC 1864
 Db 1993 AACATAGTACATAGCGTGGTGGGAGGCAATGCTTCAACCGAGGAGAGTCTTAC 1934
 Qy 1865 GAAATATGACATCTCTTACGCTGAGTACGCTTTTGGTGAATCTTGGCGGTGTGATCGGC 1924
 Db 1933 GAAATATGACATCTCTTACGCTGAGTACGCTTTTGGTGAATCTTGGCGGTGTGATCGGC 1874
 Qy 1925 TATTCGCGGATGATACCAAGCTTCTGCTGATGCTCGAATTCCTGAGGAGAGCAT 1984
 Db 1873 TATTCGCGGATGATACCAAGCTTCTGCTGATGCTCGAATTCCTGAGGAGAGCAT 1814
 Qy 1985 CCCAAGGACATTAAGGAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGATGCC 2044
 Db 1813 CCCAAGGACATTAAGGAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGATGCC 1754
 Qy 2045 GAAGTGTCTCGGCGCTCTCACTGCTTATGTACAAATTAAGTACTACAGATTCGGG 2104
 Db 1753 GAAGTGTCTCGGCGCTCTCACTGCTTATGTACAAATTAAGTACTACAGATTCGGG 1694
 Qy 2105 GAATTAAGTGGACTA----- 2121
 Db 1693 GAATTAAGTGGACTACAGTAAGCGGAACATTTCTCTCAGGTAGCGATGCACACTAA 1634
 Qy 2122 -----CAGAGTCCATCTCGATATGATGCGACAGCTACCGCGCTCANTGGGATA 2171
 Db 1633 CTGCTCTTTGAGAGTCCATCTGATATGATGCGACAGCTACCGCGCTCANTGGGATA 1574
 Qy 2172 AGGACTTCGATCTGACCTTACCTGAGGAGGCTTACACACAGAACACTGGCTTTGTGGCA 2231
 Db 1573 AGGACTTCGATCTGACCTTACCTGAGGAGGCTTACACACAGAACACTGGCTTTGTGGCA 1514
 Qy 2232 TCTATAGGTGAAGAGCGCATGATGTTCAATAGACCATCATCAGAGACCAAGGAGAA 2291
 Db 1513 TCTATAGGTGAAGAGCGCATGATGTTCAATAGACCATCATCAGAGACCAAGGAGAA 1454
 Qy 2292 CGATTCCTCCAGCAAACTTCAATTCGAGAA----- 2321
 Db 1453 CGATTCCTCCAGCAAACTTCAATTCGAGAAAGGTAGTGTACCCTCGAGCTCCCAATGAAC 1394
 Qy 2322 -----AGACTCTAAGCTGCGAGGCTCATACGAAACC 2357
 Db 1393 TTCATTACATCTGATTTCTTTGAGAACTCTAAGCGTCCGAAAGGCTCATACGAAACC 1334
 Qy 2358 GACCGGTTGTTTAAAGGAAACGAAACCTTTGAAATAAACCACCAAAAAAATAA 2412
 Db 1333 GACCGGTTGTTTAAAGGAAACGAAACCTTTGAAATAAACCACCAAAAAAATAA 1279

RESULT 4
 ADD94785

ID ADD94785 standard; cDNA; 2710 BP.

XX AC ADD94785;

XX AC ADD94785;

XX AC ADD94785;

XX AC ADD94785;

XX AC ADD94785;

XX AC ADD94785;

XX AC ADD94785;

XX AC ADD94785;

XX AC ADD94785;

XX AC ADD94785;

Mus musculus.

WO2003054008-A2.

03-JUL-2003.

18-DEC-2002; 2002WO-CA001967.

20-DEC-2001; 2001US-00028394.

(COMP-) COMPATIGENE INC.

Perreault C, McBride K;

WPI; 2003-559122/52.

P-PSDB; ADD94785;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.

Disclosure; SEQ ID NO 3; 66pp; English.

This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC)-associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestine cancer, sarcomas, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the mouse SIMP cDNA which is related to the invention. Note: this sequence does not appear in the specification but was obtained by the indexer from GenBank.

Sequence 2710 BP; 678 A; 648 C; 689 G; 695 T; 0 U; 0 Other;

Query Match 39.7%; Score 960.6; DB 10; Length 2710;
 Best Local Similarity 64.8%; Pred. No. 8, 2e-244;
 Matches 1495; Conservative 0; Mismatches 794; Indels 18; Gaps 4;

Qy 102 AGTGGCTGGCTACAGAGCCCTAATCAGCTTGGCCATCTCTCTAATCGCTGGCTGGCCG 161

Db 258 AGCGCGCGGGTGGCAGTGGTGGCTCTCTTCCATCATCTTCTGGCTGGCTGGCCG 317

Qy 162 GATTTCTCTCTCGCTTCTCGCGCTCATCCGTTTCGAGTCGATTATCCATCAGTTTCATC 221

Db 318 GCTTCAGCTCGCGCTCTTTCGCGCTCATCCGCTTCGAGAGCATCATCCAGAGTTCGACC 377

Qy 222 CGTGGTTCAACTACCGGGCCACCGCTCATCGTGCAGAGATGGTTGGTCAACTTCCTCA 281

Db 378 CGTGGTTCAACTATAGATCAACATCATCTTTCGATCTCATGAGTTTCTAA 437

Qy 282 ACTGGTTTCAGACGAGCGCATGATTCGCTCGGAGGATTGGGGGGTACCGCTCTATC 341

Db 438 ATTGGTTTGTATGAAGAGCATGGTACCCACTGGGAGAGATAGTGGTGGCAGCTTACC 497

Qy 342 CCGGCTCTGATGATACGTCCGCGGAATCCATTTGGCTCTCTGACGTACTCAACTACGG 401

Db 498 CAGGCTGTGATGATCAACAGCTGGCTTATTCATTTGATTTAAATACATTGAACATAACAG 557

Qy 402 TCCATATTCGTGACATCTCGCTGTTTCTGGGGCCCATCTTCAGTGGCTGACCTCCATCT 461

Db 558 TTCACATAGAGATGTGTGTATTCCTTTCGACCAACTTTTAGCGGCTTATACCATAT 617

source of immunodominant MHC-associated peptide; SIMP; MHC;
 major histocompatibility complex; human leukocyte antigen; HLA;
 cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 lung cancer; intestine cancer; sarcoma; prostate cancer;
 testicular cancer; breast cancer; melanomas; pancreatic cancer;
 haematological cancer; immune response; lymphoid cell proliferation;
 autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 murine; gene; ss.

KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
 KW gene; ss.
 OS Homo sapiens.
 PN WO2003054008-A2.
 XX 03-JUL-2003.
 PD 18-DEC-2002; 2002WO-CA001967.
 PF 20-DEC-2001; 2001US-00028384.
 PR (COMP-) COMPATIGENE INC.
 PA Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 XX P-PSDB; ADD94784.
 DR New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 XX acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 PT Claim 6; SEQ ID NO 1; 66bp; English.
 PS This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcoma, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;
 Query Match 39.58; Score 955.8; DB 10; Length 2481;
 Best Local Similarity 64.88; Pred. No. 1.5e-242;
 Matches 1489; Conservative 0; Mismatches 752; Indels 18; Gaps 4;
 QY 102 AGTGGCTGGGTACAGCAGCTAATACCTTCGCCATCCCTGCTAAATCGCGTGGCG 161
 DB 191 AGCGGCTGGGTGGCGAGTCGCTTCTCTCTCCATCCATCCCTCTCCGCGCTGGCTGGCG 250
 QY 162 GATTTCTCTCGCTCTCGCGTCATCCGTTTCGAGTCGATATCCATGAGTTGATC 221
 DB 251 GCTTCAGCTCGCGCTCTTTCGCGTCATCCGTTTCGAGTCGATATCCATGAGTTGATC 310
 QY 222 CGTGGTTCAACTACCGGCCACCGCTTACATGGTGCAGATGGTTGATCAACTTCCTCA 281
 DB 311 CGTGGTTCAACTACCGGCCACCGCTTACATGGTGCAGATGGTTGATCAACTTCCTCA 370
 QY 282 ACTGGTTCAACTACCGGCCACCGCTTACATGGTGCAGATGGTTGATCAACTTCCTCA 341
 DB 371 ATTGGTTGATGAAGAGCATGGTATCCACTAGAGAGATAGTAGTGGTATCTGTTTACC 430
 QY 342 CCGGCTGATGATTACCTCGCGGGGAATTCATTTGGCTGTGCACGCTACTCAACATACCGG 401

DB 431 CAGGGTTGATGATAACCGCTGGCCCTATTTCATTCGATTTTAAATACATTAACATACTG 490
 QY 402 TCATATTCGTGACATCTCGCTGTTCTCTGGCGCGGATCTTCAGTGGCTGACCTCCATCT 461
 DB 491 TTCATAAGACAGTATGTGTCTTCCCTTGCCACCACTTTTAGCGGCTTACATCTATAT 550
 QY 462 CCACCTACCTGCTGACCAAGAGCTGTGTCTCGCGGCGCGGCTCTTCGCCGCCAGCT 521
 DB 551 CTACTTCTCTGCTTACAAGAGAACTTTGGAACCAAGAGAGAGAGCTTTTAGCTGCTGTT 610
 QY 522 TCATCGCCATCGCTGCTGCTACATCAGTAGTGTGGTGGCTGATCTGATGAAGAG 581
 DB 611 TTATTCATTTACACGCTACATATCTCGGTGAGTGTGGATCTTTGATTAAGAG 670
 QY 582 GCATTGCCATATTCGCCCTGCGAGTTCACCTATCTTCTGTGGTGGCTGCTGAGTGAAGACTG 641
 DB 671 GCATTGCTATTTTGCACCTTCAGTTTCATCATACTATTTATGGGTAAATCTGTAAGAACTG 730
 QY 642 GATCGGTGCTGCTGGCGCGCGCTTTGCTCTACTTCTACATGCTGCTCGGCTGGG 701
 DB 731 GGTCACTTTTGGACAATGTGCTGCTTATCTCTATTTCTATGCTTCTGCTTGGG 790
 QY 702 GTGGTACCTGTTTCATCATCAACCTGATACCCCTGACAGTCTTCGTACTGCTCATTAAGG 761
 DB 791 GTGGTTATGTTTATCATCAATCTTATTCACCTGATGTATTTGTGTGTACTGATGC 850
 QY 762 GCAGGTACTCGCGGCTGCTGTGACCACTACGACCTTCTACATCTCTGGACTGCTGT 821
 DB 851 AGAGATACAGAAAAGAGTTCATACATAGATATAGCATTTTCTACATTTGGGTTAATAT 910
 QY 822 TCTCCATGAGATCCCTTCGTGGGATTCACCAACGATACGACCAAGTGAACATAGGCTG 881
 DB 911 TATCAATGAGATACCTTTTGGGATTCACGACCAATCAGAAAGTGAACATAGGCTG 970
 QY 882 CGCTGGAGTGTGTGCTCTTATGGCGGTGGGACCTTGGGCTTTCAGTTCGCTGCTG 941
 DB 971 CTGAGGTGCTTTCATCTGCTGAGCTTATGCTTCTTGCAGTATCTGAGAGACGAT 1030
 QY 942 TGTGCGCAACAGTTCGGAAGCTGTTTCATCTGCTGGCGGATTCGTGGTGGGCTTGGCG 1001
 DB 1031 TAACAAAACAGAGTTCAGACCTTTTCTTTTGGGTGTATCATAGTTCGAGTGTGCTG 1090
 QY 1002 TCTTTGCGCGCTGCTGCTGCTACCATGCTGGCGGTGTTGGCCCTGAGTGGAGCT 1061
 DB 1091 TGTTCCTTAGTCTCATCTATTGACTTATACAGTTACATGACCATGAGTGGAGGT 1150
 QY 1062 TCTACTCGCTGGGATTCGTGCTAGCCAAAGATCCCAATTCACATTCGATTCGCTGT 1121
 DB 1151 TTTATTCATTTGGGATTCGTGATGCAAAATACACATTTCCAAATTTATTCATCAGTGT 1210
 QY 1122 CGGAGCATCAGCCACCACTTGTGCTGCTTCTTCTTGTGCTGCAATCTGCTGTGGCG 1181
 DB 1211 CTGAGATCAACTACGACTTGGGTGCTTCTTCTTGTGCTACATATTCCTGTATGTA 1270
 QY 1182 CTTTCCAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1241
 DB 1271 CTTTCCAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1330
 QY 1242 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1301
 DB 1331 CTCTATATGCAATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1390
 QY 1302 TCACGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1361
 DB 1391 TGACTCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1450
 QY 1362 TGCAAGAGATTCGCTAAGCGAATGGGACAGCAATAGCGGACGACCAAGTGGATG 1421
 DB 1451 TGGGGATCA-----CATGAAGAGGAAATCCACCTGTGGAG-----GACAGCAGTG 1498
 QY 1422 AGCTGAGGATTCATTTAGAGAGAGACGCTGTACGACAGGCTGGGCAAGCTGAGCATC 1481

Db 1499 ATGAGGATGACAAAAGAAAACCAAGGAAATTTGTATGATAGGCGAGGTAAAGTCAGGAAAC 1558
QY 1482 GTACTAAGCATGATCCAGCAGGATCTAGCGTCAGCTCAACCTGAGAGATTTCTTA 1541
Db 1559 ATGCAACTGACAGAGAAAACCTGAAGAGGATTAGGCCCTTAATATAAAGCATTTGCA 1518
QY 1542 TTTTGGCCGTTCTAATGCTGTGTGATGATGTTGCTGTCCTCACTGCACTGCGGTGACCAACA 1601
Db 1619 CCATGTTGATGCTGATGCTATTTGATGATGTTTGTGTCCTCACTGACCTGGGTGACCAACA 1678
QY 1602 ATGCTTACTCCAGTCCCTCCATTTGCTTGGCTTTTCCACACAGTCAAGATGGATCCCGCA 1661
Db 1679 ATGCTTACTCTAGTCCAGTGTAGTCTTGGCTTATACAA---TCATGATGGACACGGA 1735
QY 1662 ACATTTAGACGATTTAGAGAGGCTTACTAGTCTTTCGCAACACACTGCCGATGATG 1721
Db 1736 ATATCTTAGATGATTTAGAGAGGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATG 1795
QY 1722 CTCGCTTATGTTCTTGGTGGATTAACGATACAGATAGCGGGAATGCAAAACAGAACGA 1781
Db 1796 CACGAGTAATGTTCTTGGTGGATTAATGCTATCAGATAGCTGGATGCTTAATAGACTA 1855
QY 1782 CGCTAGTGATTAATACGTGGAAACATATGTCATAGCGCTCGTTGGCAAGCAATGT 1841
Db 1856 CGTTGGTGATTAATACACCTTGGAAATAACAGCCACATAGCACTGGTGGGAAAAGCTATGT 1915
QY 1842 CTTCAACCGAGGAGAGTCCCTACGAAATTAATGACATCTCTTGACGTGGACTACGTTTGG 1901
Db 1916 CTTCTAATGAACACAGCGCTATAAATCANTGAGACTCTAGATGTAGATTAATGTTTGG 1975
QY 1902 TGATCTTTGGCGGTGTGATGCGCTATTTCTGCGATGATATCAACAAGTTCTCTGTTGATG 1961
Db 1976 TTAATTTTGGAGGGTTATTTGCTATTTCTGTTGATGATATCAACAATTTCTCTGTTGATG 2035
QY 1962 TCCGAAATTCCTGAGGAGAGATCCCAAGGACATTAAGGAAGCGATTAATTTACCGACC 2021
Db 2036 TTAGATAGCTGAGAGAGACATCCCAAGACATTCGGAAGTACTATTTTACCCAC 2095
QY 2022 GCGGTGAATTCAGGTAGATGCCGAAGTGTCTCGGCGCTGCTCAATGCTCTATGTATCA 2081
Db 2096 AGGAGAAATTCGCTGTAGCAAGCAGGATCCCTACTTTGTTGAATTTGCTTATGTATA 2155
QY 2082 AATTAAGTACTACAGATTCGGAATTTGAAGTTGGACTACAGAGGTCCATCTGGATATG 2141
Db 2156 AATGTCTACTACAGATTTGGAGAAATGCACTGATTTTCGTACACCCAGGTTTG 2215
QY 2142 ATGCAACAGTAACCGCGTCAATTTGGAAATAAGGACTTCGATCTGACCTACCTGAGAGG 2201
Db 2216 ACCGAACACGTAATGCTGAGATTTGGAATAAGGACATTAATTCAAACATTTGGAAGAG 2275
QY 2202 CTTACACCAACACACTGCTTGTTCGATCTATAGGTTGAAGCAAGCCGATGAGTTCA 2261
Db 2276 CTTTATCATCAGACACTGGCTTTTAGATATATAAGTAAAGCACTGATTAACAGG 2335
QY 2262 ATAGACCATCACTGAAGACCAAGGAGAACGATTCCTCCAGCAAACTTCATTTCCGAGAA 2321
Db 2336 AGACATTAGATACACAAACCTCGAGTCACCAACATTTTCCCAAAACAGAAATTTGTCAA 2395
QY 2322 AGAATCTAAGCGTCGCAAG---GGCTACATACGAAACCGACCGGTTGTTGTTAAGGAA 2378
Db 2396 AGAAGCTACCAAAAGAGAGCGTGGCTACATTAATAATAGCTGGTTTTTAAGAAAGCA 2455
QY 2379 AACGAACCTTGAATAAAC 2397
Db 2456 AGAAATATCTAAGAGAC 2474

RESULT 6

ABL02795

ID ABL02795 standard; cDNA; 2855 BP.

XX

AC

ABL02795;

XX

DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 2867.
DB Drosophila; developmental biology; cell signalling; insecticide;
KW pharmacological; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-658660/75.
DR P-PSDB; ABB58692.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 2867; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB13051), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2855 BP; 661 A; 750 C; 724 G; 720 T; 0 U; 0 Other;
Query Match 30.3%; Score 733.4; DB 4; Length 2855;
Best Local Similarity 61.0%; Pred. No. 1.5e-183;
Matches 1302; Conservative 0; Mismatches 761; Indels 72; Gaps 4;
QY 121 CCTAATCACCTTCGCCATCTCTGCTAATCGCTGCTGCGCGATTTTCTCTCGCCTCTT 180
DB 466 CTTGCTCAAGCTGGCCATCTCTCATCTGCGGCGGTTTATCATTTGCCACACGCTTGT 525
QY 181 GCGCGTCACTCGTTTGGATGATATCCATGATTTGATCGGTTGATGATGATGATGATGATG 240
DB 526 CTCTGTGCTGCGATTCGAAAGCGTAAATCCATGATTCGATCCGCTACTTCACTACCGCAC 595
QY 241 CACCGCTTACATGCTGCGAGATGTTGTTGATCACTTCTCACTTCCACAGTGGTTGATGACGCGC 300
DB 586 CACCGGTTTCTGCGGAGGAGGCGCTTTTCAAGTTTCCACAGTGGTTGATGACGCGC 645
QY 301 ATGTTATCGCTCGCAGGATTTGGCGGTACCGTCTATCCGCGCTGATGATGATGATGATG 360
DB 646 CTGTGATCCCTGGCGCGCATCTATCCGCGCGCATCTATCCGCGGTGATGATGATGATG 705
QY 361 GCGCGGAATCCATTTGGGTGCTGCGACGCTACTCAACATACCGGTCCATATTCGTGACATCTG 420
DB 706 GCGCGCTTGTACCGCTGATGCTGCTCAATGTGACCATCGACATACGAACTGTG 765
QY 421 GGTGTTCTGGCGCGATCTTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480
DB 766 CGTCTTCTGCGCGCTTCTTCTCTGCTGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825
QY 481 GAGAGTGTGTCGCGCGCGCGCTTCTTCCGCGCGCGCTTCTTCCGCGCGCGCTTCTTCCGCGCG 840

Db 826 GGAGATACACAGCACTGGAGCTGGAGCTGGTGGCGCGCTTTGATATCCATGTTCCCGG 885
Qy 541 CTACATCAGTAGGTCGGTGGCTGGATCGTAGCAATACGAGGCAATGCCATATTCGCCCT 600
Db 886 GTATATCTCTGATCCGTGGCGGATCGTAGCAATGAAGCATGCCATTTCTGCA 945
Qy 601 CGAGTTCACTTCTCTGGTGGTGGCTCAGTGAAGACTGGATCCGTCTTCTGGTGGC 660
Db 946 GCTCTTCACTTACTTATTTGGATCAAGCGGTAAAGACGGGACGATCTTTTGGTGGC 1005
Qy 661 CGAGCGCTTTGCTTACTTCTACATGCTGGCTGGCTGGTGGTGGTGGTGGTGGTGGT 720
Db 1006 TATGTCGGCATTTGGCTTACTTCTATATGCTCTCTGTTGGTGGTGGTGGTGGTGGT 1065
Qy 721 CAACCTGATACCTCTGACGCTTTGCTATGCTGCTATGATGGGAGTATCGCGCGTCT 780
Db 1066 TAACCTAATCCGCTGACGCTGCTGGCGCTGATGATACCGGACGTTTCTCGCACAGAT 1125
Qy 781 GCTGACAGCTACAGCACTTCTACATCTGGGACTGCTGTTCTCCATCCAGATCCCTT 840
Db 1126 CTACATAGCATACAGCACTTACTGCTGGCACTTCTGTCGATCGATCTGTT 1185
Qy 841 CGTGGATTCACCCGATACGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 1186 TGTGGATTCACCCGATACGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1245
Qy 901 CTTTATGCGCTGGCCACTTGGCCATTTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 1246 GTGCCAGTACAGCTTTGCTGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
Qy 961 GAAGCTGCTATCTGCGCGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 1306 TCTGCTCTTCAAGACGTTGCTTTCAGTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
Qy 1021 GCTCAGCTGCTGGCGCTTGGCGCGCTGGAGTGGAGCTTCTACTGCTGCTGGATAC 1080
Db 1366 GCTCAGCTTACCGGGAAGTCTCCCTGGACCGGAGATTTACTGCTTACTAGATCC 1425
Qy 1081 TGGTACGCAAGATCCACATTCCTATTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 1426 ATCTATGCAAGATCCACATTCCTATTCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1485
Qy 1141 TTGCTGCTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1486 CTGCTGCTTCTTATTTGATCTGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1545
Qy 1201 GTACTGATCAAGCAGATCAACGACGAGCGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 1546 TTTCTGCTTTTCCAGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1605
Qy 1261 GCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 1606 TATTTATTTCCGAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1665
Qy 1321 GCTGCGGAGTGGCTTTTGGGACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db 1666 GCTATCCGGAATGCCATTTTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1719
Qy 1381 CGGAATGGGACAGCCATAGCGAGCGGACCGGAGTGGATGAAGCTGAGGATTTCCATTGA 1440
Db 1720 -----TGCGGAGCTCATCGAAGCAGGA 1744
Qy 1441 GAAGAAGACGCTGTACGACAGGCTGCAAGCTGAGCATCTGTAAGCATGATGCCCA 1500
Db 1745 GTAGAAAGC-----AAGCGGACGACCAAGAGCTGGAACA 1779
Qy 1501 GCAGGATCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Db 1780 GCAGACGGAGCGCTGAGAGCGAGGTGGCCATTGGATTCGTGGGTGCTCATCACTCTGAT 1839
Qy 1561 GTTATGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620

Db 1840 GCTTATAGTGTACAGCTGCACTGCACTGGGTCACTCTGGAGGCTTACTCTCTGCCCCAG 1899
Qy 1621 CATTGCTTTGGCTTTCCACACAGTCAAGATGGATCCCGCAACATTTTAGACGATTTTTCAG 1680
Db 1900 TATTGTTTGGAGTGC---CAGTCCGACGATGGGGCGCATCATTTTCGATGACTTCCG 1956
Qy 1681 AGAGGCTTACTTGGCTTTCCAGAACACTGCGCATGATCTGCGTTATGCTTGGT 1740
Db 1957 CGAGGCTTACTTGGCTGAGATGAACACTCCGAGGACGCTGCAATATGCTTCTGGT 2016
Qy 1741 GGATTAACGATACCAAGATAGCGGGAATGGCAAAAGAACGACGCTAGTGGATATATATAC 1800
Db 2017 GGATTAACGCTACCAAGATAGCGGCAATGGCCATTCGACGATATTTGGTGGATACATATAC 2076
Qy 1801 GTGGAAATAGTGTACATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGGAGATC 1860
Db 2077 TTGGAAACACACATATATCGCGCTTGGCGAGCGGATGGCTTCTTCGGAGGAGAAAGC 2136
Qy 1861 CTACGAAATTAATGACATCTTGTACGCTGACTAGCTTTTGGTGGATCTTTGGCGGTGAT 1920
Db 2137 CTACGAGATATAGGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 2196
Qy 1921 CGCTTATCTGGGATGATATCAACAGTTCCTGCTGGATGCTGCAATGCTGAGG--- 1977
Db 2197 TGCTTACTCATCGGACGATATCAACAGTTCCTGCTGGATGCTGCAATGCTGAGG--- 2256
Qy 1978 AGAGCATCCCAAGGACATTAAGGAAAGGATTTACTTTACCGACCGCGTGAATTCAGGGT 2037
Db 2257 GGATCGTGGTGGCACTCCCGGAAAGGACTTACTATCGGCCAACGAGAGTTCGAGT 2116
Qy 2038 AGATCGGAGGCTGCTCGGCTGCTCACTGCTTATGTAACAATTAAGCTTACTACAG 2097
Db 2317 GGAACAGGAGGCTCACCACACTGCTCAATTTGTTGATGTAAGATGCTGCTACTATCG 2376
Qy 2098 ATTTCGGGAAATGAAGTTGGACTACAGAGGCTCATCTGGATATGATCGCACACGTAACGC 2157
Db 2377 CTTTGGCAATGTACAGGAGGTTGCAAGCCGAGGCTACGATCGAGTTCGTGCGC 2436
Qy 2158 GGTATGGATTAAGGACTTGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2217
Db 2437 CGAGATCGGCAACAGGACTTTGAAGTGTGATGAGGAGGCTGATGAGGAGGCTGATGAGGAGGCT 2496
Qy 2218 CTGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2252
Db 2497 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2531

RESULT 7

ABL18224/c

ID ABL18224 standard; DNA; 2953 BP.

XX

AC ABL18224;

XX

DF 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6145.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

XX

KW pharmaceutical; gene; ds.

XX

OS Drosophila melanogaster.

XX

BN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

XX

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 6145; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention of
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
CC AB572072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2953 BP; 924 A; 628 C; 648 G; 753 T; 0 U; 0 Other;

Query Match 28.8%; Score 696; DB 4; Length 2953;
Best Local Similarity 100.0%; Pred. No. 1.3e-173;
Matches 696; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAAGCGAAGATGTGTGTTGCAATTCAGATCGGTTATATATTTTCGAGTTACTGCTG 60
DB 696 TCTAAGCGAAGATGTGTGTTGCAATTCAGATCGGTTATATATTTTCGAGTTACTGCTG 637
QY 61 GAATGGGACATGATCGGACCGGACGATCTGACAGCAAGTGGCTGGCTACACAG 120
DB 636 GAATGGGACATGATCGGACCGGACGATCTGACAGCAAGTGGCTGGCTACACAG 577
QY 121 CCTAATCACTTCGGCATCTCTAATCGCTGGTGGCGGATTTTCCTCTCGCCCTTT 180
DB 576 CTAATCACTTCGGCATCTCTAATCGCTGGTGGCGGATTTTCCTCTCGCCCTTT 517
QY 181 CGCGTCATCCGTTGAGTGTGATTCATGAGTTGATCCGTTGTTCACTACCGGCG 240
DB 516 CGCGTCATCCGTTGAGTGTGATTCATGAGTTGATCCGTTGTTCACTACCGGCG 457
QY 241 CACCGCTACATGGTGCAGAAATGGTGTGTAACATTCCTCAACTGGTTTCAGCAGCGCGC 300
DB 456 CACCGCTACATGGTGCAGAAATGGTGTGTAACATTCCTCAACTGGTTTCAGCAGCGCGC 397
QY 301 ATGGTATCCGCTCGCAGGATGTGGGCGTACCGTCTATCCGGCGCTGATGATGCTC 360
DB 396 ATGGTATCCGCTCGCAGGATGTGGGCGTACCGTCTATCCGGCGCTGATGATGCTC 337
QY 361 CGCGGGAATCCATTTGGCTGCTGACGTACTCAACATACCGGTCCATATTCGTGACATCTG 420
DB 336 CGCGGGAATCCATTTGGCTGCTGACGTACTCAACATACCGGTCCATATTCGTGACATCTG 277
QY 421 CGTGTTCCTGGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGTCACCA 480
DB 276 CGTGTTCCTGGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGTCACCA 217
QY 481 GGAGCTGTGGTCCGGCGGCGCGGCTCTTTCGCGCCAGCTTCATCGCCATCGGCTCGG 540
DB 216 GGAGCTGTGGTCCGGCGGCGCGGCTCTTTCGCGCCAGCTTCATCGCCATCGGCTCGG 157
QY 541 CTACATCAGTAGTGGTGGCTGGATCTGATGATGACAGGCGATTCGCATATTCGCGCT 600
DB 156 CTACATCAGTAGTGGTGGCTGGATCTGATGATGACAGGCGATTCGCATATTCGCGCT 97
QY 601 GCAGTTCACTACTTCTCTGCGGTGGCTTCAGTGAAGACTGGATCCGTTCTGGTGGCG 660
DB 96 GCAGTTCACTACTTCTCTGCGGTGGCTTCAGTGAAGACTGGATCCGTTCTGGTGGCG 37
QY 661 CGCAGCGCTTTGCTCTACTTCTACATGTTGTCGCG 696

Db 36 CGCAGCGCTTTGCTCTACTTCTACATGTTGTCGCG 1

RESULT 8

ABT20789
ID ABT20789 standard; DNA; 2232 BP.
XX
XX
AC ABT20789;
XX
XX 16-APR-2003 (first entry)
XX
XX Aspergillus fumigatus essential gene #3147.
XX
XX Fungicide; cytosolic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response; ds.
XX
XX Aspergillus fumigatus.
OS
XX WO200286090-A2.
PN
XX 31-OCT-2002.
XX
XX 23-APR-2002; 2002WO-US013142.
XX
XX 23-APR-2001; 2001US-0285697P.
PR
XX 27-APR-2001; 2001US-0287066P.
PR
XX 05-JUN-2001; 2001US-0295890P.
PR
XX 09-JUL-2001; 2001US-0303899P.
PR
XX 31-AUG-2001; 2001US-0316362P.
PR
XX (ELIT-) ELITRA PHARM INC.
XX
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
PI
XX WPI; 2003-093124/08.
XX
XX New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
PT
XX
PS Disclosure; Page; 175pp; English.
XX
XX The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterization, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This polynucleotide sequence represents one of the essential
CC genes of Aspergillus fumigatus of the invention
XX
SQ Sequence 2232 BP; 473 A; 605 C; 545 G; 609 T; 0 U; 0 Other;

Query Match 26.3%; Score 635.4; DB 8; Length 2232;
Best Local Similarity 58.9%; Pred. No. 1.4e-157;

		Matches 1248; Conservative 0; Mismatches 796; Indels 75; Gaps 6;			
QY	137	ATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGGCTCTTGGCGTCACTCGTTTC	196		
DB	76	ATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGGCTCTTGGCGTCACTCGTTTC	135		
QY	197	GAGTCGATTTATCCATGATTTGATCCGCTGTTCAACTACCGGCGCACCGCTACATGTTG	256		
DB	136	GAGAGTATATCCACGAGTTGACCGCTGTTCAACTTCGAGCAACAAATACATAGTA	195		
QY	257	CAGATGTTGCTACATCTCTCACTGTTTCGAGCGGCGCATGTTATCGCTCGGC	316		
DB	196	CAGATGTTGCTACATCTCTCACTGTTTCGAGCGGCGCATGTTATCGCTCGGC	255		
QY	317	AGGATTTGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGGGAATCCATTGG	376		
DB	256	CGTGTACCGGTGACGTTATATCCCGGTCTCATGGTACGAGCGGCTGATACGAT	315		
QY	377	CTGCTGACGTAATCAACATACCGTCCATATTCGTGACATCTGGTGTTCGTGGCGG	436		
DB	316	ATCTTGGGATTTCTACTATCCCGGTGCAATTCGCAACATCTGGCTCTACTGGCGCA	375		
QY	437	ATCTTCACTGCGCTGACCTCCATCTCCACCTACTCTGACCAAGGAGCTGTGGTCCGG	496		
DB	376	GGATTTCCCGGCTGACGTTGCAATGGCAATGATCTGTGACATCCGAGATGTCTCTCG	435		
QY	497	GGC---GCCGCTCTTGGCGGCGAGCTTCATCGCCATGTCCTGCTGACATGATAGG	553		
DB	436	CCATCTGAGGCTTCTTGGAGAGCTTTTCATGGGAATCGCCCTGGTACATCTCCGGA	495		
QY	554	TCGGTGGCTGATCGTACGTAACGAGGCAATTCGCCATATTCGCCCTGACGTTCACTAC	613		
DB	496	TCAGTGTGGAAGCTACGTAACGAGGCAATTCGCCATATTCGCCCTGACGTTCACTAC	555		
QY	614	TTCTGTGGGTGCGCTCAGTGAAGACTGATCGGTGTTCTGGTGGCGCGACGCTTTG	673		
DB	556	TTCTGTGGGTGCGCTCAGTGAAGACTGATCGGTGTTCTGGTGGCGCGACGCTTTG	615		
QY	674	TCCTACTTTACATCTCGCTCGGCTGGGTGGCTAGCTGTTCATCATCACTGATACCC	733		
DB	616	TTCTACGCTACATGTTGTCGGATGGGTGGGTATGTTCTTCATACGAACTGATCCCG	675		
QY	734	CTGACGCTTCTGTAATCTATGCGGAGACTGATCGCGCTGCTGCTGACGACTAC	793		
DB	676	CTGACGCTTCTGTAATCTATGCGGAGTAGATAGACACTGCGATCTACATTTAGCTAT	735		
QY	794	AGCACCTTTACATCTCGGAGCTGCTTCTCCATGACAGATCCCTTCGTGGGATCCAA	853		
DB	736	ACCACATGTTATGGCTGGGACTTTGGCTAGCATGCAGATCCCTTCGTGGATTTTG	795		
QY	854	CCGATAGCACGAGTGAACATGCTGCTGGGAGTGTGCTGCTCTTATGCGGTG	913		
DB	796	CCTATCCGAACACGACCACTGCTCGCTTGGGTGCTTTCGCGCTGCTTCAGCTTGTG	855		
QY	914	GCCACCTTGGCCATTTGCACTGCTGCTGCGCAACGAGTTCGGAAGCTGTTCATC	973		
DB	856	GCTTTCGCGAGTTTGTCCGAGCTTCGTTCCAAAGCAAGCAGTTCAGAGACTTCGTACC	915		
QY	974	GTCCGCGATTTGCTGGGCTGGGTGCTTGTGGCGCTGCTGCTGCTACCATGCTG	1033		
DB	916	GCCATGATTTTACATCTGCTGCTGGTTCGTTGGCTGCTGCTGCTGCTGCTGCTG	975		
QY	1034	GGCGTTGGCCCGCTGGAGTGAAGCTTCTACTGCTGTGGGATACGCTGACCCAAAG	1093		
DB	976	GGAGTGTGCTGCTGCTGGAGCGGCGATTTCTACTCTGTGGGACACTGGCTATGCCAAA	1035		
QY	1094	ATCCACATTTCCCATTTGATCGTTCGAGGATCAGCCACCACTTGTTCGTCTC	1153		
DB	1036	ATCCACATTTCCCATTTGATCGTTCGAGGATCAGCCACCACTTGTTCGTCTC	1095		
QY	1154	TTCTTTGATCTGCAATCTCTGGTGTGCGCTTCCCAAGTGGAGTGTGGTACTGCATCAAG	1213		
DB	1096	TTCTTTGATCTGCAATCTCTGGTGTGCGCTTCCCAAGTGGAGTGTGGTACTGCATCAAG	1155		

QY	1214	CAGATCAACGACGAGCGGTTTTTCGTGCTGTAGCCATCAGTGGGTTTACTTCCGT	1273		
DB	1156	GACCTCAAGAGAGAGATGTTCTTCATTAATCTACTCGTCTTGGAGCTACTTCCGC	1215		
QY	1274	GCTGTGATGCTGCTTGTGATGTTGACCTCAACGCGGTGTTGATGCTGCTGGCGAGT	1333		
DB	1216	GCTGTGATGCTGCTTGTGATGTTGACCTCAACGCGGTGTTGATGCTGCTGGCGAGT	1275		
QY	1334	GCCTTTTCGGGACTGTTGATGTTTCTCTCAAGAGGATTCTTAAGCGAATGGGCACA	1393		
DB	1276	GCCTGTCTTCATTTCTCGACACCTATAG-----	1305		
QY	1394	GCATTAAGCGGACGACCAAGTGGATGAAGCTGAGGATTCATTGAGAGAGAGACGCTG	1453		
DB	1306	GGACTACCTCCCGACACCAAGCGTCTGAAG-----G	1338		
QY	1454	TACGACAGGCTGGCAAGCTGAAGCATCTACTAAGCATGATCCCGAGGATACCTGGC	1513		
DB	1339	AAACGATGAAGATCTGCTTTCACACCTCTTCCTCAGTTCGAAAGCCCAATTTGGA	1398		
QY	1514	GTGAGCTCCAACTG---AAGAGTATGTTATTTTGGCGCTTCTAATGCTGTGATGATG	1570		
DB	1399	ATCACCTCCATGTTTCTAAGATTATAGTGAAGCGCTCTGTTGCTGCTACTCTCCTG	1458		
QY	1571	TTGCTGTCCACTGACGCTGGGTGACCAAGTACTCTCAGTCCCTCCATTTGCTTCTG	1630		
DB	1459	TTGTTGGCATGTCACCTGGGTACATCGAATGCTACTCTTCTCTTCCGTTGCTTCTG	1518		
QY	1631	GCTTTCCCAACAGTCAAGATGATGATCCCGCAACATTTTACGATTTTACAGAGGCTTAC	1690		
DB	1519	GCTAGTTCGATGCT---GACGGAAGCAATACATCATTCAGGATTCCTGAGGCTTAC	1575		
QY	1691	TACTGCTTTCCGACACACTGCGCATGATGCTCGCTTATGTTCTGGTGGATTTACGGA	1750		
DB	1576	TACTGCTTTCCGACACACTGCGCATGATGCTCGCTTATGTTCTGGTGGATTTACGGA	1635		
QY	1751	TACGATAGCGGGAATGGCAACAGAACAGAGCTGATGATGATGATGATGATGATGATG	1810		
DB	1636	TATCAATCGGTGGATGGCGACCGCCAACTTGGTTGACAAACAACACCTGGAAACA	1695		
QY	1811	AGTCACATAGCTGTTGGGCAAGCAATGCTTCAACCGAGGAGAGTCTCTACGAATT	1870		
DB	1696	ACCATATTTGCTACGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATG	1755		
QY	1871	ATGACATCTTTGAGTGGACTACGTTTGGTGGATGATGATGATGATGATGATGATGATG	1930		
DB	1756	CTCCGCGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1815		
QY	1931	GGCGATGATCAACAAAGTTCCTGTTGATGATGATGATGATGATGATGATGATGATGATG	1990		
DB	1816	GGCGATGATCAACAAAGTTCCTGTTGATGATGATGATGATGATGATGATGATGATGATG	1875		
QY	1991	GACATTAAGAAAGCGATTTACTTTACGACCGGCTGATTTACGGGTAGATGCGGAGGT	2050		
DB	1876	GAGGTTAAAGAGCGGACTTCTTTACTGACCGGCTGATTTACGGGTAGATGCGGAGGT	1935		
QY	2051	GCTCGGCGCTCTCACTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG	2110		
DB	1936	ACCCAACTATGGCAACAGCTTGTGTTGATGATGATGATGATGATGATGATGATGATGATG	1986		
QY	2111	AAGTTGGAATCAGAGGCTCCATCTGATGATGATGATGATGATGATGATGATGATGATGATG	2170		
DB	1987	AACCTCTCTTCCCGTGGGCGCAAGCTGTCGACCGGCTGCGTGGGTCAAAACTTCCACA	2046		
QY	2171	AAGGACTTCGATCTGACCTTACCTGAGGAGGCTTACACACAGAACACTGCTGCTTCTCG	2230		
DB	2047	GAAGGCTCTGATCTCTACATCTGAGAGCTTTCACGAGGAGACTGATCATTCGT	2106		
QY	2231	ATCTAGGCTGAAGAGC	2249		
DB	2107	ATCTAGGCTGAAGAGC	2125		

	Query Match	26.2%	Score 632.8	DB 10	Length 2472
	Best Local Similarity	58.2%	Prod. No. 6.9e-157		
	Matches 1247	Conservative 0	Mismatches 802	Indels 93	Gaps 4
QY	118	CAGCGTAATACACCTTCGCCCATCTCGTAAATCGCGCTGGCTGGCGCGAATTTCTCTCGGCT	177		
Db	151	CACACTTTTGAAGTCTCTCAATCTGTCAATGGTGTGTATATATCCTTCTCATCTCGTCT	210		

us-10-028-384-7.rng

Thu Dec 16 16:25:13 2004

1291 CAGCATGACATTTTTCAGCTGTAATGTCGCTGCTAATGCTAGTGTGGCACTGTTATGAG 1350
 1318 CATGCTGCGGAGTGGCTTTTGGGACTGTTGGATGTTCTTCTGCAAGAGGATTCGTC 1377
 1351 CATTCTCTGTCATGAGTCTCCAGTCTG----- 1384
 1378 TAAGCGAATGGGCACAGCCATAAGCGCCAGCCAGCGAAGTGGATGAAGCTGAGGATTCAT 1437
 1385 -----TCCAC 1389
 1438 TGAAGAAGACGCTGTACGCAAGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGC 1497
 1390 ATACATGAAGATCTGGACATAGTTCGCCCGCAGCAAGAAGAGCAAGAGCAAGATTC 1449
 1498 CCAGCAGGATACCTGGCGTCAGCTCCCAAGCTGAAGAGTATTGTTATTTTGGCCGTTCTAAT 1557
 1450 CACCTACCTATTAGATT-----GAAGTGCAGTGGATGATCTGCTCATGGCTTT 1503
 1558 GCTGTGATGATGTTGCTGTCCACTGACGTGGGTGACCGACATGCTACTCCAGTCC 1617
 1504 CTTTCTCATCACTACACTTTTCACTCACTGGGTGACCGATGAGGCTACTCTTCTCC 1563
 1618 CTCCTATGCTTGGCTTTCCCAACAGCTCAAGATGATCCCGCAACATTTTAGACGATT 1677
 1564 GTCCATTGACTATCTGCC-----GTGGTGGATGCGAGTAGGATCATATTGATGACTT 1620
 1678 CAGAGAGGTTACTACTGCTTTCCGAGAACACTCGCGGATGATGCTCGGTATGCTTTG 1737
 1621 CCAGAGCATATATTGCTTCTGTCATTAATCTCCAGAGGATGCGAAGGTCATGCTCG 1680
 1738 GTGGATACGATACCATAGCGGATGCAACAGACGACGCTAGTGGATAATA 1797
 1681 GTGGATTATGGCTATCAGTTACAGTATGCAATGCAACCAATTTAGTGGCAATAA 1740
 1798 TAGTGGAAATAGTACATAGCGCTGTTGGCAAGCAATGTTTCAACCGAGGAGAA 1857
 1741 CACATGGAATAATACCATATTCTCGATAGGAGGCAATGGGCTCCACAGAGGAAAA 1800
 1858 GTCTAGCAAAATATACATCTTTGACGTGACGCTGTTTGGTGGATTTTGGCGGT 1917
 1801 AGCTATGATCATGAGGAGCTCGATGTACGATGCTGCTGCTGCTATTTTGGAGGCT 1860
 1918 GATCGGCTATCTGGCGATGATATCAACAGTTTCTGTGGATGCTCGCAATTCGTGAGG 1977
 1861 CACTGGGATTTCTCTGATGATATCAACAGTTTCTTGGATGCTCGGATTCGAGGAG 1920
 1978 ---AGAGCATCCCAAGACATTAAGAAAGCATTACTTTACCGACCGGCTGAATTCAG 2034
 1921 CACAGATACAGGCAACATATCAAGAGAAATGACTATTATATCTCCAACTGGGGAGTCCG 1980
 2035 GGTAGATGCGAAGGTGCTCCGCGCTGCTCACTGCTTATGTAACAAATTAAGCTACTA 2094
 1981 TGTGACCGTGAAGTTCTCCAGTGTCTCACTGCTCACTGCTCACTGCTTACTA 2040
 2095 CAGATTGGGAAATGAAGTTGAGTACAGAGTCCATCTGATATGATCGACACAGTAA 2154
 2041 TCGCTTTGGACAGGTTTACACAGAACCCAGCGCTCTCCAGGCTTTGACCGGTGTCGAAA 2100
 2155 CGCGTCTATTGGGAATAAGGACTTCGATCTGACCTACCTGGAGAGGCTTACACACAGA 2214
 2101 TGTGAGATTGGGATTAAGACTTTGAGTTGATGCTTGGAGAGGCTTATACACAGA 2160
 2215 AACTGCTGTTTGGCATCTATTAGGTTGAAGAGCGCATGA 2256
 2161 ACATTGGCTGTCAGGATATACAGGTAAAGGACCTGGATAA 2202

RESULT 10
 ADH28844
 ID ADH28844 standard; DNA; 2472 BP.
 XX
 AC ADH28844;

XX 11-MAR-2004 (first entry)
 XX Human chronic myelogenous leukaemia (CML) gene marker #112.
 DE
 XX de; chronic phase chronic myelogenous leukaemia; CP-CML;
 KW blast crisis CML; BC-CML; human; chronic myelogenous leukaemia;
 KW gene marker.
 XX Homo sapiens.
 OS
 XX US2003104426-A1.
 PN
 XX 05-JUN-2003.
 PD
 XX 14-JUN-2002; 2002US-00171581.
 PF
 XX 18-JUN-2001; 2001US-0298914P.
 PR
 XX (LINS/) LINSLEY P S.
 PA (MAOM/) MAO M.
 PA (DAIH/) DAI H.
 PA (HEYI/) HE Y.
 PA (RADI/) RADICH J P.
 XX Linsley PS, Mac-M, Dai H, He Y, Radich JP;
 PI WPI; 2003-787046/74.
 XX
 DR Classifying cell sample as chronic phase chronic myelogenous leukemia or
 XX blast crisis chronic myelogenous leukemia by detecting difference in
 PT expression of genes corresponding to the markers such as X15415, U89436.
 PT
 XX Disclosure; SEQ ID NO 112; 31pp; English.
 PS
 XX The invention relates to a method of classifying a cell sample as chronic
 CC phase chronic myelogenous leukemia (CP-CML) or blast crisis CML (BC-
 CC CML). The method is useful for classifying a sample as CP-CML or BC-CML.
 CC The present sequence represents a human chronic myelogenous leukaemia
 CC (CML) gene marker used to distinguish blast crisis CML from chronic phase
 CC CML.
 XX
 XX Sequence 2472 BP; 566 A; 568 C; 583 G; 755 T; 0 U; 0 Other;
 SQ
 Query Match 26.2%; Score 632.8; DB 10; Length 2472;
 Best Local Similarity 58.2%; Pred. No. 6.9e-157;
 Matches 1247; Conservative 0; Mismatches 802; Indels 93; Gaps 4;
 QY 118 CAGCCTAATCACCTTCGCAATCCTGCTAATCGCTGGCTGGCGGATTTTCTCTCGCT 177
 Db 151 CACATTTTGAAGTCTCTCATTTCTGTCATGCTGCTGTTATTCCTTCCACTCGTCT 210
 QY 178 CTTGCGCGTCATCCGTTTCGAGTCGATATCCATGAGTTTGTATCCGTTTCACTACCG 237
 Db 211 GTTTCGTCTCGTGAAGTTTGAAGTGTATCCATGAGTTTGTATCCGTTTAAATTATCG 270
 QY 238 GGCACACCGCTACATGTCGAGAAATGTTGGTCAAACTTCCTCAACTGGTTCCAGCAGCG 297
 Db 271 GACTACCAAGTTTCTGCTGAGAGGGTTTATTAATTCCTAATCTGTTGATGACCG 330
 QY 298 CGCATGTTATCCGCTCGGACGAGTATGTTGGCGGTACCGTCTATCCCGGCTGATGATTAC 357
 Db 331 AGCTGTGTACCTTTGGACGAATCATTTGGAGGAACAATTTTCCAGGTTTAAATGATCAC 390
 QY 358 GTCCGGGGATCCATTTGGCTGTCGACGATCTCAACATACCGTCCATATTCGTGACAT 417
 Db 391 CTCGCTGCAATCTACCATGTTCTCCATTTTCCCATCATCACCATCGACATTCGATGT 450
 QY 418 CTGGCTGTTCTGGCGCGGATCTTTCAGTGGCTGACCTCCATCTCCACCTTACTGCTGAC 477
 Db 451 CTGTGTGTTCTGGCCCT 510
 QY 478 CAAGAGCTGTGTTCGCGGGCGGCGCTCTTCGCGCGGCGGCTTTCATCGCCATCGTGCC 537

Db 511 CAAGAGCTCAAGGATGAGGGGCTGGCTTCTTGTCTGCCATGATCTCTAGTTC 570
Qy 538 TGGCTACATCAGTAGTGGTGGTGGATCGTAGATAAGAGGAGGATGCCATATTCG 597
Db 571 TGGATATATCTCCGATCTGGCTGGCTCTATGATAATGAGGATGCCATCTTTG 630
Qy 598 CTGAGCTACCTACTCTCTGGTGGTGGCTGAGTGAAGACTGGATCGTCTGGTC 657
Db 631 CATGTACTCACTACTACTAGTGGATCAAGGAGTAAAGACTGGTTCATCTGGG 690
Qy 658 GCGCCAGCCGCTTGTCTCTACTACTATCATGGTCTCGCCCTGGGGTGGCTACGTTCAT 717
Db 691 AGCTAAGTGGCCCTTGTCTATTTCTACATGGTCTCGTCACTGGGAGGTTATGTTCCT 750
Qy 718 CATCAACTGATACCCCTGACGCTCTGCTACTCTCTATGATGGGAGTACTCGCGG 777
Db 751 GATCAACTTAATCTCTCCAGCTCTGCTGATGCTCACAGCCGCTTCTCTCACCG 810
Qy 778 TCTGCTGACCACTACAGCACTTCTACATCTCTGGGACTGCTGTCTCCATGAGATCCC 837
Db 811 GATCTATGTGGCTACTGCTGTTTACTGCTGGTACTATATCTTTCTAGGAGATCTC 870
Qy 838 CTTCGTGGATTCGACGATACGACAGTGAACATGCTGCTGCTGGAGTGTGT 897
Db 871 CTTCGTGGTTCGAGCTGCTCTTCTATCAGACATGCGAGGCTTGGGTCTTGG 930
Qy 898 GCTCCTTATGGCCGTGGCCACTTTCGCGCATTTGCACTGCTGCTGCGGCAACAGTT 957
Db 931 TCTCTGCCAGATCCATGCTTGTGGATTACTCTGCGCAGCAAGTTGAATCCACAAT 990
Qy 958 CCGAAGCTGTATCTGCTGGGATTCGCTGGGATTCGCTGGGCTGCTTGTGCGCTGT 1017
Db 991 TGAAGTCTTTTCCGAGGCTCATCTCTGCTGGAGCTTGTCTCTCTACCTGGGAGC 1050
Qy 1018 GGTGCTCACCATGCTGGGCTGTGGCCCGCTGGAGTGGAGCTTCTACTCGCTGGGA 1077
Db 1051 TCTCCTCATGCTGACAGGAAATATCTCCTGACGCGGCTTCTACTCACTGTGGA 1110
Qy 1078 TACTGGCTACGCAAGATCCACATCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1137
Db 1111 TCCCTCTATGTAAGAACACATCCCATCACTGCTGCTGCTGCTGCTGCTGCTGCTG 1170
Qy 1138 CACTTGGTCTCTCTTCTTGTATCTGACATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1197
Db 1171 AACCTGGTCTCATACTATTTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1230
Qy 1198 GTGCTACTGATCAGAGATCAAGCAGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCT 1257
Db 1231 CTATTACTGCTTTAGCAACCTGCTGATGCGCGGATTTTATCATGATGATGCTGCTG 1290
Qy 1258 TCGGCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
Db 1291 CAGCATGATCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1350
Qy 1318 CATGCTGCGGAGTGGCTTTTGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377
Db 1351 CATTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1384
Qy 1378 TAAGGAAATGGGCACAGCCATAAGCGCACCCAGAGTGGATGAGCTGAGGATTCAT 1437
Db 1385 -----TCCAC 1389
Qy 1438 TGAGAGAGAGCTGTAGCAGAGCTGGAGCTGAGCATCGTACTAGCATGATGC 1497
Db 1390 ATACATGAAGATCTGGACATAGTGGCCAGACAGAGAGCAAGAGCATTC 1449
Qy 1498 CCAGCAGGATCTGGGCTGAGCTCCCACTCAAGAGTATGTTATTTTGGCCGCTTCTAAT 1557
Db 1450 CACCTACCTATTAAGATT-----GAAGTGGCAAGTGGGATGATCTGCTGCTGCT 1503
Qy 1558 GCTGTGTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617

Db 1504 CTTTCTCATCACCTACACCTTTTCATTTCAACCTGGTGACAGAGGAGGCTACTCTCTCC 1563
Qy 1618 CTCCATTGTCTTGGCTTTCCACAACTAGTCAAGATGGATCCCGCAACATTTTAGACGATTT 1677
Db 1564 GTCCATTGTATCT 1620
Qy 1678 CAGAGAGCTTACTACTGCTTTTCCGAGAACACTGCGGATGATGCTGCGGCTTATGCTTGT 1737
Db 1621 CCGAAGACATATTTTGGCTTCTGCTATATATCTCCAGAGATGCGAAGTCTATGCTCTG 1680
Qy 1738 GTGGATTACGGATACAGATAGCGGATGGCAACAGACAGACGCTAGTGGATTAATAA 1797
Db 1681 GTGGGATTTATGGCTATCAGATTTACAGCTATGGCAACCGAACAAATTTTAGTGACAATAA 1740
Qy 1798 TACGTGAAACAATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1857
Db 1741 CACATGAAATATACCCATATTTCTCGAGTAGGAGGAGGATGCTGCTCAGAGAGGAAA 1800
Qy 1858 GTCTACGAAATATAGATCTCTTGGAGTGGACTAGCTTTTGGTGTATCTTTGGCGGTGT 1917
Db 1801 AGCCTATGAGATCATAGGAGGCTCGATGTCAGCTATGCTGCTGCTCATTTTGGAGGCT 1860
Qy 1918 GATCGCTATTTCTGGGATGATATCAACAAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1977
Db 1861 CACTGGGTATCTCTCTGATGATCAACAATTTCTTGGATGCTGCGATTTGGAGGAG 1920
Qy 1978 ---AGAGCATCCCAGGACATTAAGGAAAGCATTTTACCGACCGGCTGAATTCAG 2034
Db 1921 CACAGATACAGGCAACATATCAAGGAGATGACTATTACTCCAACTGGGAGTTCG 1980
Qy 2035 GGTAGATCGCGAGGCTCTCGGCTCTCAACTGCTTATCTATCAAAATTAAGTACTA 2094
Db 1981 TGTGACCGTGAAGTCTTCCAGTCTGCTCACTGCTCATGTAAGATGTTACTA 2040
Qy 2095 CAGATTGGGAAATGAAGTTGAGTACAGAGTTCATCTGATATGATGCTACAGCTAA 2154
Db 2041 TCGCTTTGGACAGCTTTACACAGAGCCAGCGCTCTCCAGGCTTTGACCGTCTCGGAA 2100
Qy 2155 CGCGTCTATGGGAAATGAAGTTCGATCTGACTCTGAGAGGAGGCTTACACACAGA 2214
Db 2101 TGCTGAGNTGGGAAATGAAGCTTTGAGCTGATGCTCTGAGAGAGGCTATACACAGA 2160
Qy 2215 ACATGCTGCTTCTGCTGCTATAGGCTGAAGAGCGGCTATGA 2256
Db 2161 ACATGCTGCTGCTGCTGCTATACAGGTTAAAGGACCTGGATAA 2202

RESULT 11

ADD94791

ID ADD94791 standard; DNA; 3094 BP.

XX AC ADD94791;

XX AC ADD94791;

DT 29-JAN-2004 (first entry)

XX XX

DE Mouse ITM1 gene sequence.

XX XX

KW source of immunodominant MHC-associated peptide; SIMP; MHC;

KW major histocompatibility complex; human leukocyte antigen; HLA;

KW cytotoxic; immunosuppressive; antineoplastic; gene therapy; cancer;

KW lung cancer; intestine cancer; sarcoma; prostate cancer;

KW testicular cancer; breast cancer; melanoma; pancreatic cancer;

KW haematological cancer; immune response; lymphoid cell proliferation;

KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;

XX murine; gene; ds; ITM1.

OS Mus musculus.

XX XX

XX WO2003054008-A2.

XX EN

XX 03-JUL-2003.

XX PD

XX 18-DEC-2002; 2002WO-CA001967.

XX PF

XX	20-DEC-2001; 2001US-00028384.	
XX	(COMP-) COMPATIGENE INC.	
XX	Perreault C, McBride K;	
XX	WPI: 2003-559122/52.	
DR	P-PSDB; ADD94792.	
XX		
XX	New human source of immunodominant MHC-associated peptide (SIMP) nucleic	
PT	acids and proteins, useful for diagnosing and treating cancers, e.g. lung	
PT	or breast cancer, or for suppressing an immune response in an autoimmune	
PT	disease.	
XX		
XX	Disclosure; SEQ ID NO 9; 66pp; English.	
XX		
CC	This invention relates to a novel isolated or purified human protein,	
CC	termed source of immunodominant major histocompatibility complex (MHC) -	
CC	associated peptide (SIMP), which is expressed ubiquitously in human	
CC	cells, where the protein has the potential of generating several protein	
CC	fragments binding with high affinity to a human leukocyte antigen (HLA)	
CC	molecule. The invention may allow development of therapeutics with	
CC	cytostatic or immunosuppressive activity or provide sequences useful for	
CC	antisense therapy or gene therapy. The source of immunodominant MHC-	
CC	associated peptide (SIMP) nucleic acids, proteins and fragments are	
CC	useful for diagnosing and treating cancers, for example lung cancer,	
CC	intestinal cancer, sarcomas, prostate cancer, testicular cancer, breast	
CC	cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP	
CC	proteins are also useful for modulating an immune response. Decreasing	
CC	lymphoid cell proliferation is useful for suppressing an immune response	
CC	responsible for an autoimmune disease or a transplant rejection. The	
CC	present sequence is that of the mouse Itm1 gene which is related to the	
CC	invention. Note: This sequence does not appear in the specification but	
CC	was obtained by the indexer from GenBank.	
XX		
XX	Sequence 3094 BP; 758 A; 651 C; 703 G; 982 T; 0 U; 0 Other;	
QQ		
Query Match	26.1%; Score 631.2; DB 10; Length 3094;	
Best Local Similarity	58.2%; Pred. No. 2.1e-156;	
Matches 1246; Conservative	0; Mismatches 803; Indels 93; Gaps 4	
QY	118 CAGCCTAATCACCTTCGGCATCTGCTAATCGCTGCTGCGCGGATTTTCCTCTCGCT 177	
Db	156 CACACTTCAAGAGCTTCTCATCCCTGCGATGGCTGTGTATATCTTTCTACTCGTCT 215	
QY	178 CTTCCGCGTATCCGTTTCGAGTGCATATCCATGAGTTTGATCCGCTGTTCAACTACCG 237	
Db	216 TTTTCTGTGTGCAGATTGTAAAGTGCATCCATGAGTTTGATCCGCTACTTAAATTATCG 275	
QY	238 GGCACACGCCCTACATGGTGCAGATGTTGTGTACACTTCCTCACTGGTTTGACGAGCG 297	
Db	276 GACTACCCGGTTTCTGGCTGAGGAGGGGTTTTAAATTCCAATCTGTTTGTATGACCG 335	
QY	298 CGCATGGTATCCGCTCCGACAGATTTGGGCGGTACCGTCTATCCGCGCTCATGATTAC 357	
Db	336 GCCTTGGTACACCTTTGGGCCGAATCAITGGAGGAACAATTTACCAGGTTTAATGATCAC 395	
QY	358 GTCCGCGCGGAATCCATTGGCTGCTGCAGGTACTCAACATACCGTCCATATTCGTGACAT 417	
Db	396 TTCTGCTGCATCTACATGTACTCCATTTCTTCATTCACATATTGACATTCGGATGT 455	
QY	418 CTGCGTGTTCCTGGCGCGATCTTTCAGTGGCCTGAOCTCCATCTCCACCTTACTCTGTGAC 477	
Db	456 CTGTGTTTTCCTGGCCCCACCTTTTCTCCTCTTTTACCACCATCGTTTACGTACCACTTAC 515	
QY	478 CAAGGAGTGTGTTCCGCGGGCGCGGCTCTTCGCGCGCAGCTTCATCGCCATCGTGGC 537	
Db	516 CAAAGAGCTCAAGATGACAGAGCTGGGCTCTTGTGTGCCCATGATTCGTGTAGTTC 575	
QY	538 TGCGTACATCAGTAGGTCCGTGGCTGGATGTCAGATACAGGGGCAATTCGCATATTCGC 597	
Db	576 TGGGTATATTTTCGATCTGTAGCTGGCTCTATGATTAATGAAGGAATTTGATCTTTTG 635	

1006	CTTGTA	CAGCTGATCGGATTCGTGCGAAGTCGTCCGACGACTCGTGTGCTCTGGCAAGCAATTC	1065
959	CGAAGCTGTTTCATCGTCGCGGGAATGCTGTGTGGCGGTGTGGCGTCTTTGTGTGGCGCTGCTG	1018	
1056	CAGCTCCTTTCTCAAAGAGCTTTTGTGCTGTGGCGGTATTCTGCCTCAGTTTGTGCTGCCTCGTC	1125	
1019	GTGCTACCAATGCTGTGGCGTTGTGGCCCGTGGAGTGGACGCTTCTACTCGCTGTGGGAT	1078	
1126	ACTTTGACTTTCTCTGATGGATTCGCCCTTCCTCGCTGGAAATTTTATTCTTTTGGGAT	1185	
1079	ACTGCTACGCCAAGATPCCACATTTCCCATCATTTGCATCCGTGTGCGAGCATACGCCCAACC	1138	
1186	ACTGGCTATGCGAAGGTCCACATGCCCATTTATTGCTCCGTCTCGAAACACCAAGCCCAACC	1245	
1139	ACTTGGTCTCGTCTCTTTTGTATCTGCACATCTCGTGTGGCGCTTCCCAAGTGGAGATG	1198	
1246	GCTTGGCGCTCATTTCTACTTTTGACCTCGAAATGCTTATCTTCTTTTCCCTGCGGTGTC	1305	
1199	TGGTACTGCATCAAGACAGATCAACACGACGCGCTTTTCGTGTGTGTATAGGCATCAGT	1255	
1306	TTCTGGTGTTTCAAGAGAGCTTCCGATGACGAGATCTTTCATCATCATTTATATGCGGTCTTC	1365	
1259	CGCGTTTACTTCGCTGTGTGATGTGTGTCGTTGATGTTGACCCCTCACGCGGTGGTGTGC	1318	
1366	AGTGCCTATTTTGGCGTGTCTAGTTTCGACTTATGCTTTGTATCATCGCTGTGTCTGT	1425	
1319	ATGCTGCGCGAGTGGCTTTTTCGGGACTGTGTGGATGTGTTCTTGCAAGAGGATTCGTCT	1378	
1426	GTTTCTCTCCGCCATTGGTTTCTCCAACTTCTCGAGGCGTATATGTGACCCCGCTATCCGCC	1485	
1379	AAGCGAATGGGCACAGCCATAAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTCATTT	1438	
1486	GAAAGCAGCAGGAGAGTGTGCGAGTCTCAGACGCGAGTTGTCTCCAGTCCAGGGCGAG	1545	
1439	GAGAAGAGAGCGCTGTACGACAGGCTGGCAAGCTGAAGCATCTGATTAAGCATGATGCC	1498	
1546	AAGATGGCGCTGCACACGCCAATAAGACGGGTTCTCTTTCACAGGATTTTGTAGCGGC	1605	
1499	CAGCAGGATACCTGGCGTCAGCTCCAACTCGAAGAGTATTGTTATTTTGGCGCTTTCAATG	1558	
1606	AAGTCTGCTCCGGCATCTTTGGTCTCGACATCGATTTGCTGTGGTTCCATCTCTCT	1665	
1559	CTGTTGATGATGTTCCGTGTC---GACTCGACGTGGGTGACCGAATGCTCTACTCAGT	1615	
1666	GTCCTCTCTTCATCTTTGCTCTACTGCAATATGTGACTTCAAAGCGATTTCTTCTG	1725	
1616	CCCTCCATTTGCTTGGCTTTCCACACAGCTCAAGATGGATCCCGCAACATTTTAGACGAT	1675	
1726	CCTTCAGTGGTACTTGCATCGCGAATC---CCGGATGTTAGCGAAATATCATTTGATGAT	1782	
1676	TTGACAGAGGCTTACTACTGGCTTTTCGAGAACACTGCGGATGATGCTCGGTATGTCT	1735	
1783	TTCCGAGAGGCTTACTACTGATTCGCCAAAACACCGCCGAAGACACGGTCATCATGTCC	1842	
1736	TGTTGGGATTAACGATACAGATAGCGGGAATGGCAAAACAGAAACGCTAGTGGATAT	1795	
1843	TGTTGGGATTAACGCTACCGATACCGATCGCTGTGTATGGCTGATCGCCCAACCTTTGATAAC	1902	
1796	AATACGTGGACATATGTCATAGCGCTGTGTCAGAGGCAATGCTTCAAACGAGGAG	1855	
1903	AATACCTTGGAAATACACCCACATTCGACAGTTGTGTAGGCCATGGCTTTCAAACGAAT	1962	
1856	AAGTCTTAACGAAATATGACATCTTTGACGTGGACTACGTTTTTGTGTGATCTTTTGGCGGT	1915	
1963	GTCGCATATCTTATCTTGGAGAGCATGATGTGATTTACGTTCTGTGTGATCTTTGGGGC	2022	
1916	GTGATCGCTATTTCTGCGGATGATATCAACAAGTTCTCTGTGATGGTTCGGAATTCGTAG	1975	
2023	TTATTGGGCTACTCTGTGTGACATATCAACAAGTTTTTGTGTGATGTTGATCTCACAA	2082	
1976	CGAGAGCATCCCAAGGACATTAAGGAAACGGATTTACTTTACCGACCGCGGTGAATTCAGG	2035	
2083	GGTGAATGGCGCTGACGAGGTGCGAGGATCAACTACTTTACTCAAGAGGGGAGTATGCT	2142	

2036	GTAGATGCCGAAGGTGCTCCGGCCCTGCTCAACTGCCTTATGTACAAATTAAGCTACTAC	2095
QY		
2143	GTCCGATGACAGGGCCACCCCTACTATGAAGAACTCTCTCATGTACAAATGTCTTACTAC	2202
DB		
2096	AGATTCCGGGAATTGAAGTTGGACTACAGAGGTCCTCTGGATATGATCGCACACGTTAC	2155
QY		
2203	CGCTTCCCCGA-----GCTTTATGGTGGACACCCGGCTCAAGACAGGGTTCGAGGC	2253
DB		
2156	GCGGTCAATTGGGAATAAGCACTTCGATCTGCACCTACCTGGAGGAGGCTACACCAGAA	2215
QY		
2254	CAAAATTATCCCTCCTTAACAGTGTTACTCTTGATACTCTTGACGAAGGGTTCATCCGNA	2313
DB		
2216	CAC TGCTGTTTGGCATCTATAGGGTGAAGAAG	2248
QY		
2314	AAATTGATGTCAGGATCTTACAAGTCAAGAAG	2346
DB		

RESULT 13
ABV24502
ID: ABV24502 standard; cDNA; 2760 BP.
XX
XX
ABV24502;
AC
AC
XX
DT
16-SEP-2002 (first entry)
XX
XX
Human prostate expression marker cDNA 24493.

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

Claim-1; Page 4632-4633; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (AEU0010-ASV6221); of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) also useful as a pharmacodynamic or pharmacogenomic marker.

Sequence: 2760 BP; 688 A; 645 C; 658 G; 769 T; 0 U; 0 Other;

1

20

Query Match 25.5%; Score 615.6; DB 5; Length 2760;
Best Local Similarity 58.3%; Pred. No. 2.7e-152;
Matches 1250; Conservative 0; Mismatches 799; Indels 95; Gaps 6;

QY	118	CHAGCCTAATACCTTCGGCATCTCTGCTAATCGCCTGGCTGGCGGGATTTTCCTCTCGCCT	177
DB	149	CACACTTTTGAAGCTTCTCATCTGCTCAATGGCTGCTGTATATTCCTTCTCCACTCGTCT	208
QY	178	-CTTGGCGGTATCCGTTTCAGTGCATATATCCATGATTTGATCCCGTGTCAACTACC	236
DB	209	GTTTTGTCTGCTGAGATTTGAAGTGATTCATCCATGATTTGATCCGCTACTTAAATATC	268
QY	237	GGCCACCCGCTACATGTGTGAGAAATGTTGGTCAACTTCCCTCACTGTTGCGAGCAGC	296
DB	269	GGACTACCAAGTTCCTGCTGAGGAGGGGTTTAAATTCATTAATTCATTAATGATGACC	328
QY	297	GGCCATGATATCCGCTCGGCGAGGATTTGGCGGCTACCGTCTATCCCGGCTCATATTA	356
DB	329	GAGCCTGGTACCTTTGGGAGCAATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCA	388
QY	357	CGTCCGGGGATCATTTGGCTGCTGACGTA-CTCAACATACCGGTCCATATTCGTGAC	415
DB	389	CTCTGCTGCAATCATACCATGATCTCCATTTTTCACATCACCATCGCATTCGGAAT	448
QY	416	ATCTGGGTGTTCTCGCGCCGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTG	475
DB	449	GTCTGTGTTCCTGGCCCTCTCTTCTCCTCTTCCACCATCGTCACTACCACTT	508
QY	476	ACCAAGGAGCTGTGTTCGGGGGCGCGCTCTTGGCGGCGAGCTTCATCGCGATCGTG	535
DB	509	ACCAAGAGCTCAAGGATGCAAGGGCTGGGCTTCTGCTGCTGCATGATTTGCTAGTT	568
QY	536	CTGGCTACATCAGTAGTGGCTGGATCGATACGATAACGAGGCGATTCGCAATTC	595
DB	569	CTGGATATATCTCCGATCTGTGGCTGCTCTATGATAATGAAGGGATTCGCACTTT	628
QY	596	GCCTGCGAGTCTCTGTGGTGGCTCAGTGAAGACTGATCCGTGTTCTGTG	655
DB	629	TGCATGCTACTCACCTACTATGATGGATCAAGGAGTAAAGACTGTTCCATCTGTGG	688
QY	656	TGGCGCGAGCGCTTGTCTACTTCTACATGCTGCTCGGCTGGGCTGCTACGTTGTC	715
DB	689	GCAGCTAAGTGCGCTTGTCTTATTTCTACATGCTCGTCAATGGGAGGTTATGTTTC	748
QY	716	ATCATCACTGATACCCCTGACGCTTCTGCTACTGCTCATATTAAGGCGAGTACTCGCG	775
DB	749	CTGATCAACTTAATTCCTCCACGCTCTGCTGCTGATGCTCAAGGCGCTTCTCTCAC	808
QY	776	CGTCTGCTGACAGCTACAGACCTTCTACATCTCGGACTGCTGTTCTCCATCGAGATC	835
DB	809	CGATCTATGTGGCTACTGCTGTTTACTGCTTGGCCTGCGCACTATCTTTCTATGCAATC	868
QY	836	CCCTTCGTGGATTCGAACCGATACGACCGAGTGAACAATGCTGCTGGCTGGAGTGT	895
DB	869	TCCTTGTGGGTTTCCAGCTGTCTTTCATCAGAGCACATGGCAGCTTGGGGCTTT	928
QY	896	GTGCTCCCTATGCGCGTGGCCACCTTGGCGCATTTGCACTGCTGCTGCGCAACGAG	955
DB	929	GGTCTCTGCCAGATCATGCTCTTCTGATTTACCTGGCGCAAGTTGATTCACACAA	988
QY	956	TTCCGAGAGCTTTCATGCTGGCGGATGCTGCTGGCGGTGGCGTCTTGTGGCGCTC	1015
DB	989	TTTGAAGTCTTTTCCGAGGCTCATCTCTCTGTAGGCTTGTCTTCTCACCGTGGGA	1048
QY	1016	GTGGTGTCTCACCATGCTGGGGTGTGGCCCGCTGGAGTGAGCGCTTCTACTCGCTGTGG	1075
DB	1049	GCTCTCTCATGCTGACAGGAAATATCTCTCTGGACGGGCTTCTACTCGCTGCTG	1108
QY	1076	GATATGCTACCGCAAGATCCATTCCTCATTTGATTCGCTGGGAGTCAAGCC	1135
DB	1109	GATCCCTCTTATGCTTAAGAACAAATCCCATCTATGCTTCTGTGTCTGAGCATCAGCC	1168
QY	1136	ACCACTTGTTCTCTGCTTCTTTTGTATCTGCACATCTGGTGTGGCGCTTCCAGTGGGA	1195

DB	1169	ACAACCTGGTCTCTCATACTATTTTGACCTGCAGCTCTCTCTCTCATGTTTCCAGTGGC	1228
QY	1196	GTGTGCTACTCATCAAGCAGATCAACGACGAGGCGTTCCTGTGTGCTGTGACGCAATC	1255
DB	1229	CTCTATTACTCTTTAGCAACCTGCTGATGCGCCGATTTTATCATCATGATGTGTG	1288
QY	1256	AGTGGGTTTACTTTCGGTGTGATGGTGTGATTTGATTTGATCCCTCACGCGGTGGTG	1315
DB	1289	ACCAGCATGTACTTTTCAGCTGTAATGCTGGCTTAATGCTAGTGTGGCACCTGTATG	1348
QY	1316	TGCATGCTGGCGGAGTGGCTTTTCGGACTGTTGATGTGTTCTGCAAGAGGATTCG	1375
DB	1349	TGCATTTCTCTGTCGATTTGGAGTCTCCAGGTGCTG-----	1384
QY	1376	TCTAAGCGAATGGGCGACACCATTAAGCGGACCCACGAGTGAAGCTGAGGATTC	1435
DB	1385	-----TCC	1387
QY	1436	ATTGAGAAGAAGACGCTGTACGACAGGCTTGGGAGCTGAAGCATCGTACTAAGCATGAT	1495
DB	1388	ACATACATGAAGAATCTGGACATAAGTCTCCAGACAAGAAAGAGCAAGCAACAGGAT	1447
QY	1496	CCCCAGCAGGATACCTGGCGTCAGCTCCAACTGGAAGTATTGTTATTTTGGCGCTTCTA	1555
DB	1448	TCCACTTACCCTA-----TTAAGATGAAGTGCAAGTGGGATGATCTGGTCAATGGCT	1501
QY	1556	ATGCTGTGTGATGTTGGCTTCCACTGCGCTGGGTGACCAAGCATGCTTACTCAGT	1615
DB	1502	TTCCTTCTCATCCTACCTTTTCACTTCACTGGGTGACCACTGAGGCTTACTTTCT	1561
QY	1616	CCCTCCATTTGCTTGGCTTTCCAAACAGTCAAGATGGATCCCGCAACATTTTACAGCAT	1675
DB	1562	CGTCCATTTGCTGTCTGCC---GTGGTGGGATGGCAGTAGGATCATATTTGATGAC	1618
QY	1676	TTCCAGAGGCTTACTTACTGCTTTCGACAACTGCCCATGATGCTCGCTTATGCT	1735
DB	1619	TTCCGAGAGCATATTTTGGCTTCTGCTAATACTCCAGAGGATGCAAGGCTCATGTCC	1678
QY	1736	TGGTGGATTAACGATACCAAGTAGCGGGAATGGCAACAGACGAGCTAGTGGATAAT	1795
DB	1679	TGGTGGATTAATGCTATCAGATTTACAGCTATGGCAACCGAACAATTTTATGSGACAAT	1738
QY	1796	AATAGTGGAAACATAGTACATAGCCTGTTGGCAAGGCAATGTTCAACCGAGGAG	1855
DB	1739	AACACATGGATAATATCCCATATTTCTCGAGTAGGCGAGCAATGGCTCCACAGAGAA	1798
QY	1856	AAGTCTTACGAAATTAATGACATCTTTCAGCTGGAATGCTTTCGCTGATCTTTGGCGGT	1915
DB	1799	AAAGCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTCATTTTGGAGC	1858
QY	1916	GTGATCGGCTATTTCTGGGATGATATCAACAAGTTCCTGTGGATGGTCCGAATGCTGAG	1975
DB	1859	CTCATGCGTATTCCTCTGATGATATCAACAAGTTCCTTTGGATGGTCCGATTTGAGGG	1918
QY	1976	GG---AGAGCATCCCAAGGACATTAAGGAAAGGATTTTACCGACCGCGTGAATTC	2032
DB	1919	AGCAGATACAGCAACATATCAAGGAGATGACTATATATCTCAACTGGGAGTTC	1978
QY	2033	AGGTAGATCCGGAAGTGTCTCGGCCCTGCTCACTGCTTATGTTACAAATTAAGCTAC	2092
DB	1979	CGTGTGACCGTGAAGGTTCTTCAGTGTCTCTCACTGCTCATGTACAGATGTGTAC	2038
QY	2093	TACAGATTTGGGGAAATGAAGTTGGACTACAGAGGTCCTCTGATATGATCCACACGT	2152
DB	2039	TATCGTTTGGACAGGTTTACACAGAACCCAGCGTCTCCAGCTTTTACCGTGTCCGA	2098
QY	2153	AACCGCTCATTTGGGATAAGGCTTGGATCTGACCTCACTGCTTATGTTACAAATTAAGCTAC	2212
DB	2099	AATCTGAGATTTGGGATAAGGCTTTTGGCTTGTGCTGAGGAGCATATACCA	2158
QY	2213	CAACACTGGCTGTCTCCATCTATAGGTTGAAGAGCCGATGCA 2256	

23.3%; Query Match
59.8%; Score 562; DB 8; Length 1848;
Best Local Similarity
Pred. No. 3.6e-138;

QY 1377 CTAAGCAATGGGACACGACCAATAGCGCAGCCAGGAGTGGATGAGCTGAGGATCCA 1436
Db 1129 -----GGGACTACCTCCCGACACCGCGTCTGAAGC----- 1160
QY 1437 TTGAGAAGAGACGGCTGTAGCAGACGGCTGGCAGCTGAGCAGCTGTAAGCATGATG 1496
Db 1161 -----GAAACGAAATGAAGACTCGTCTTCAACACCTCTTTCGTCAGTTC 1204
QY 1497 CCGAGCAGATGACTGGCGTCAAGCTCCAACT--GAAGAGATATGTTATTTGGCCGTT 1553
Db 1205 GGAAGCCCAATGTTGGAATCACTCCCATGTTCTAAGATTATAGTGACGCGCTGTG 1264
QY 1554 TAATGCTGTGATGATGCTCGTCTCACTGCTGCTGACGAGCAAGTATGCTTCTCA 1613
Db 1265 TCGTCTACTGCTGCTGTTGTTGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1324
QY 1614 GTCCCTCCATGCTTGGCTTTCCACACAGTCAAGATGATGATCCCGCAACATTTAGACG 1673
Db 1325 CTCCTTCCTGCTTGGCTAGTGGATGCT--GACGGAAGCCAATACATCATTTGACG 1381
QY 1674 ATTTCAGAGAGCTTACTGCTTTCGAGAACACTGCGGATGATGCTCGGTTATGT 1733
Db 1382 ATTATCGTGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1441
QY 1734 CTGTGTGGATTTACGATACAGATAGCGGAATGGCAACAGAACAGCAGCTAGTGATA 1793
Db 1442 CATGTGGATTTATGGTATCAATATCGTGGCATGGCGACCGCCCAACCTTGGTGACA 1501
QY 1794 ATAATACGTGGAGATAGTACATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGG 1853
Db 1502 ACAACACCTGGAAACACCAATATGTTACGGTGGTAAAGGCGATGAGTCAACGGAGG 1561
QY 1854 AGAAGTCTTACGAAATTTATGATCTCTTACGCTGGACTAGTTTGGTGAATCTTTGGCG 1913
Db 1562 AAGTCAGCTACCCATCTCCCGCAGCATGATGTCGATTACGCTGCTGCTGCTGCTG 1621
QY 1914 GTGTATGCTGCTTCTGGGATGATATCAACAGATTCCTGATGATGCTGATGCTG 1973
Db 1622 GTCTGCTAGTATTTCTGGGATGATATCAACAAATCTTATGATGCTGCTGCTGCTG 1681
QY 1974 AGGAGAGATCCCAAGAGCATTAAAGAAAGCGATTACTTTACCGACCGCGGTGAATCA 2033
Db 1682 AAGGTATCTGGCCGATGAGGTTAAAGAGCGGACTTCTTATGTCACGCGGTGAATC 1741
QY 2034 GGGTAGATCCCAAGGCTGCTCGGCGCTGCTCAACTGCTTATGATCAAAAT 2085
Db 1742 GTGTGACAGATGGAGCGACCCCAACTATGTCGCAACAGCTTATGATGATGAT 1793

RESULT 15

ADD94787

XX ADD94787 standard; DNA; 2733 BP.

XX AC ADD94787;

XX DT 29-JAN-2004 (first entry)

XX DE Yeast STT3 gene sequence.

XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;

XX KW major histocompatibility complex; human leukocyte antigen; HLA;

XX KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;

XX KW lung cancer; intestinal cancer; sarcoma; prostate cancer;

XX KW testicular cancer; breast cancer; melanomas; pancreatic cancer;

XX KW haematological cancer; immune response; lymphoid cell proliferation;

XX KW autoimmune disease; transplant rejection; SIMP-derived peptide; yeast;

XX KW gene; ds; STT3.

XX OS Saccharomyces cerevisiae.

XX PN MO2003054008-A2.

XX XX

PD 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

XX 20-DEC-2001; 2001US-00028384.

XX (COMP-) COMPATIGENE INC.

XX Perreault C, McBride K;

XX WPI; 2003-559122/52.

XX P-PSDB; ADD94788.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

18-DEC-2002; 2002WO-CA001967.
20-DEC-2001; 2001US-00028384.
(COMP-) COMPATIGENE INC.
Perreault C, McBride K;
WPI; 2003-559122/52.
P-PSDB; ADD94788.
New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancer, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.
Disclosure; SEQ ID NO 5; 66pp; English.
This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC) - associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestinal cancer, sarcomas, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the yeast STT3 gene which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from Genbank.

XX Sequence 2733 BP; 712 A; 597 C; 583 G; 841 T; 0 U; 0 Other;

Query Match 21.6%; Score 521.8; DB 10; Length 2733;

Best Local Similarity 54.8%; Pred. No. 2e-127;

Matches 1211; Conservative 0; Mismatches 877; Indels 123; Gaps 4;

QY 105 TGGCTGGCTACAGCAGCTTAATCACCTTCGCCATCTGCTAATCGCTGGCTGGCGGAT 164

Db 244 TGCTGTGTTTCAGACCATCTCAAGCTCGTCACTTCGTGGCGATTTTGGGGCTGCCA 303

QY 165 TTTCCTCTCGGCTCTTGGCGCTCATCGTTTCGAGTCGATTATCCATGATTTGATCCCGT 224

Db 304 TATCATCAGTTTGTTCAGTCATCAAAATTCGAGTCTATTATCAATGATTCGACCCCT 363

QY 225 GGTCAACTACGGGCGCACCGCTACATGTGAGATGTTGTTGATCACTTCTCACT 284

Db 364 GGTCAATATATAGGGCTACCAAAATATCTCGTCAACAAATTCGTTTCAAGATTTTGAAT 423

QY 285 GGTTCGACGAGCGGCATGTTATCCGCTCGCGAGGATTTGGCGGATTTTGGCGGCT 344

Db 424 GGTTCGACGAGCGGTACCTGTTACCTCCCTCGAGGGTTACTGGAGGACTTTATATCTG 483

QY 345 GCCTGATGATTTACGTCGGCGGAAATC---CATTTGGCTGTCGACGACTCAATACCGG 401

Db 484 GTTTCATGACACTAGTGTGCGTTTATCTGGCAGCGCCCTGGCGAACTGGTTGGGTTGCCA 543

QY 402 TCCATATTCGTGACATCTGGTGTCTCTGGCGCGGATCTTCACTGGCGCTGACCTCCATCT 461

Db 544 TTGACATAGAAACGTTTGTGTCTATTTCGCCCACTATTTCTGGGGTCAACCGCTGGG 603

QY 462 CCACCTACTGCTGACCAAGAGCTGTGGTCGGGGCGCGGCTCTTCGCCGCGAGCT 521

Db 604 CGACTTACGAATTTACGAAAGAGATTAAAGATCCAGCGCTGGCTTTTGGCTGCTGGT 663

1028	DB	GT	TTTGGTGGATCTTTGGCGGTGTGATCGCGCTATTCTCGCGATGATATCAACAAAGTTCCCTG	1087
1955	QY	TG	ATGTCGCGAATTGCTGAGGAGAGCATCCCAAGGACATTAAAGAAAGCGATTACTTT	2014
1088	DB	TG	ATGTCGCGAATTGCTGAGGAGAGCATCCCAAGGACATTAAAGAAAGCGATTACTTT	1147
2015	QY	ACC	GACCGCGTGAAATTCAGGGGTAGATCGCAAGGTGCTCGGCCCTGTCAACTGCCTT	2074
1148	DB	ACC	GACCGCGTGAAATTCAGGGGTAGATCGCAAGGTGCTCGGCCCTGTCAACTGCCTT	1207
2075	QY	ATG	TACAAATTAAGCTTACTACAGATTTCGGGGNATTGAAGTTGAGCTACAGAGTCCATCT	2134
1208	DB	ATG	TACAAATTAAGCTTACTACAGATTTCGGGGNATTGAAGTTGAGCTACAGAGTCCATCT	1267
2135	QY	GG	ATATGATCGCACACGTAACCGCTCATTTGGGAATAAGGACTTCGATCTGACCTACCTG	2194
1268	DB	GG	ATATGATCGCACACGTAACCGCTCATTTGGGAATAAGGACTTCGATCTGACCTACCTG	1327
2195	QY	GAG	GAGCGCTTACACCCACAGAACACTGGCTTGTTCGCCATCTATAGGGTGAAGAAGCGCAT	2254
1328	DB	GAG	GAGCGCTTACACCCACAGAACACTGGCTTGTTCGCCATCTATAGGGTGAAGAAGCGCAT	1387
2255	QY	GAG	TCTAATAGCCATCACTGGAAGACCAAGGAGAGACGATTCCTCCAGCAACTTCATT	2314
1388	DB	GAG	TCTAATAGCCATCACTGGAAGACCAAGGAGAGACGATTCCTCCAGCAACTTCATT	1447
2315	QY	TCG	AGAAAGAACTCTTAAGCGTCGAAAGGGCTACATACGAAACCGACCGGTTGTGTTAAG	2374
1448	DB	TCG	AGAAAGAACTCTTAAGCGTCGAAAGGGCTACATACGAAACCGACCGGTTGTGTTAAG	1507
2375	QY	GG	AAAAACGAACCTTGAAATAAAACCCAAAAAATAAAAAA	2412
1508	DB	GG	AAAAACGAACCTTGAAATAAAACCCAAAAAATAAAAAA	1545

US-09-270-767-158

```

RESULT 3
US-09-270-767-15440
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

```

Query Match	31.3%;	Score 756.4;	DB 4;	Length 1660;
Best Local Similarity	99.9%;	Pred. No. 8.7e-202;		
Matches 757;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1655	TCCGCGAACATTTTACACGATTTACAGAGCGCTTACTGGCTTTTCGCGAGAACCTGCC	1714	
Db	788	TCCGCGAACATTTTACAGCGATTTTCAGAGAGCGCTTACTGGCTTTTCGCGAGAACCTGCC	847	
QY	1715	GATGATGCTCGCGTTATGTCTTTGGTGGATTACGGATACCAGATAGCGGGAATGGCAAC	1774	
Db	848	GATGATGCTCGCGTTATGTCTTTGGTGGATTACGGATACCAGATAGCGGGAATGGCAAC	907	
QY	1775	AGAACGACGCTAGTGGATTAATAATACGTGGAAACAATAGTCACATAGCGCTGGTTGGCAAG	1834	
Db	908	AGAACGACGCTAGTGGATTAATAATACGTGGAAACAATAGTCACATAGCGCTGGTTGGCAAG	967	
QY	1835	GCAATGCTTTCAACCGAGGAGAAGTCTTACGAAATTATGACATCTCTTGACGTGGAGCTAC	1894	
Db	968	GCAATGCTTTCAACCGAGGAGAAGTCTTACGAAATTATGACATCTCTTGACGTGGAGCTAC	1027	


```

QY 1895 GTTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGATATCAACAAGTTCTGT 1954
Db 1028 GTTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGATATCAACAAGTTCTGT 1087
QY 1955 TGGATGGTCCGAATGTCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTTT 2014
Db 1088 TGGATGGTCCGAATGTCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTTT 1147
QY 2015 ACCGACCGCGGTGAATTCAGGGTAGATGCGCAAGGTGCTCCGGCCCTGCTCAACTGCGTT 2074
Db 1148 ACCGACCGCGGTGAATTCAGGGTAGATGCGCAAGGTGCTCCGGCCCTGCTCAACTGCGTT 1207
QY 2075 ATGTACAAATTAAGCTACTACAGATTCCGGGAATTTGAAGTTGGACTACAGAGGTCCATCT 2134
Db 1208 ATGTACAAATTAAGCTACTACAGATTCCGGGAATTTGAAGTTGGACTACAGAGGTCCATCT 1267
QY 2135 GGATATGATCCGACACGCTGATCCCGGTCTATTTGGGAATTAAGGACTTCGATCTGACCTG 2194
Db 1268 GGATATGATCCGACACGCTGATCCCGGTCTATTTGGGAATTAAGGACTTCGATCTGACCTG 1327
QY 2195 GAGGAGGCTACACACAGAACACTGGCTTGTTCGCATCTATAGGTTGAAGAGCCGCAT 2254
Db 1328 GAGGAGGCTACACACAGAACACTGGCTTGTTCGCATCTATAGGTTGAAGAGCCGCAT 1387
QY 2255 GAGTTCAATAGACCATCACTGAAGACCAAGAGAGAACGATTCTCTCCAGCAAACTTCAIT 2314
Db 1388 GAGTTCAATAGACCATCACTGAAGACCAAGAGAGAACGATTCTCTCCAGCAAACTTCAIT 1447
QY 2315 TCGAGAAGAACTCTTAAGCGTCCGAGGCTACATAGCAAAACCGGCTGTGTTGTAAG 2374
Db 1448 TCGAGAAGAACTCTTAAGCGTCCGAGGCTACATAGCAAAACCGGCTGTGTTGTAAG 1507
QY 2375 GGAACAGAACTTGAATAAACCACCAAAAAA 2412
Db 1508 GGAACAGAACTTGAATAAACCACCAAAAAAATAA 1545

```

RESULT 4

```

US-09-270-767-158/c
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

```

```

Query Match 26.6%; Score 643.2; DB 4; Length 1660;
Best Local Similarity 97.3%; Pred. No. 4.8e-170;
Matches 654; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1742 GATTACGGATACAGATAGCGGAATGGCAACAGACGCTAGTGGATAATAATACG 1801
Db 811 GAAATCGTCTAAATGTTCGGGAATGGCAACAGACGCTAGTGGATAATAATACG 752
QY 1802 TGGAAACAATAGTCACATAGCGGTGGTGGCAAGGCAATCTTCAACCGAGGAGAGTCC 1861
Db 751 TGGAAACAATAGTCACATAGCGGTGGTGGCAAGGCAATCTTCAACCGAGGAGAGTCC 692
QY 1862 TAGCAAAATTAATGACATCTCTTGACGTGGACTAGCTTTTGGTATCTTTGGCGGTGTGATC 1921
Db 691 TAGCAAAATTAATGACATCTCTTGACGTGGACTAGCTTTTGGTATCTTTGGCGGTGTGATC 632
QY 1922 GCGTATCTCGCGATGATATCAACAAGTTCTGTGGATGGTCCGAATTTGCTGAGGAGAG 1981

```

```

Db 631 GGCTATTCGGCGATGATATCAACAAGTTCTCTGTGGATGGTCCGAATTTGCCGAGGAGAG 572
QY 1982 CATCCCAAGACATTAAGGAAAGCGATTACTTTTACCGACCGGTGAATTTCAAGGTAGAT 2041
Db 571 CATCCCAAGACATTAAGGAAAGCGATTACTTTTACCGACCGGTGAATTTCAAGGTAGAT 512
QY 2042 GCCGAAGGTGTCCGGCCCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTC 2101
Db 511 GCCGAAGGTGTCCGGCCCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTC 452
QY 2102 GGGGAATTCAGTTGAGCTACAGAGTCCATCTCGATATGATCGCACAGCTAACCGCTC 2161
Db 451 GGGGAATTCAGTTGAGCTACAGAGTCCATCTCGATATGATCGCACAGCTAACCGCTC 392
QY 2162 ATTGGGAATTAAGGACTTCGATCTGACCTACCTGGAGGAGGCTCACACACAGAACTCTGG 2221
Db 391 ATTGGGAATTAAGGACTTCGATCTGACCTACCTGGAGGAGGCTCACACACAGAACTCTGG 332
QY 2222 CTTGTTCCGATCTATAGGTTGAAGAGCGGCTAGTTCAATAGACCATCTCTGAAGACC 2281
Db 331 CTTGTTCCGATCTATAGTGTGAAGAGCGGCTAGTTCAATAGACCATCTCTGAAGACC 272
QY 2282 AAGGAGAGAAACGATTCTCCAGCAAACTTCAATTCGAGAAAGAACTCTAAGCGTCCGAG 2341
Db 271 AAGGAGAGAAACGATTCTCCAGCAAACTTCAATTCGAGAAAGAACTCTAAGCGTCCGAG 212
QY 2342 GCTACATACGAAACCGGCTGTTTGTAAAGGAAACGAACTTGAATAAACCACCA 2401
Db 211 GCTACATACGAAACCGGCTGTTTGTAAAGGAAACGAACTTGAATAAACCACCA 152
QY 2402 AAAAAA 2413
Db 151 AAAAAAATAA 140

```

RESULT 5

```

US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

```

```

Query Match 26.6%; Score 643.2; DB 4; Length 1660;
Best Local Similarity 97.3%; Pred. No. 4.8e-170;
Matches 654; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1742 GATTACGGATACAGATAGCGGAATGGCAACAGACGCTAGTGGATAATAATACG 1801
Db 811 GAAATCGTCTAAATGTTCGGGAATGGCAACAGACGCTAGTGGATAATAATACG 752
QY 1802 TGGAAACAATAGTCACATAGCGGTGGTGGCAAGGCAATGTCTTCAACCGAGGAGAGTCC 1861
Db 751 TGGAAACAATAGTCACATAGCGGTGGTGGCAAGGCAATGTCTTCAACCGAGGAGAGTCC 692
QY 1862 TAGCAAAATTAATGACATCTCTTGACGTGGACTAGCTTTTGGTATCTTTGGCGGTGTGATC 1921
Db 691 TAGCAAAATTAATGACATCTCTTGACGTGGACTAGCTTTTGGTATCTTTGGCGGTGTGATC 632
QY 1922 GCGTATCTCGCGATGATATCAACAAGTTCTGTGGATGGTCCGAATTTGCTGAGGAGAG 1981
Db 631 GCGTATCTCGCGATGATATCAACAAGTTCTGTGGATGGTCCGAATTTGCCGAGGAGAG 572

```

QY 1982 CATCCAGGACATTAAGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGCTAGAT 2041
DB 571 CATCCAGGACATTAAGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGCTAGAT 512
QY 2042 GCCGAAGGTGCTCCGGCCCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTC 2101
DB 511 GCCGAAGGTGCTCCGGCCCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTC 452
QY 2102 GGGGAATTGAAGTTGGATCAAGAGTTCATCTGGATATGATCGACACAGTAAACCCGTC 2161
DB 451 GGGGAATTGAAGTTGGATCAAGAGTTCATCTGGATATGATCGACACAGTAAACCCGTC 392
QY 2162 ATTGGGAATAGGACTTCGATCTGACCTACTCGAGGAGGCTTACACACAGAACTCTGG 2221
DB 391 ATTGGGAATAGGACTTCGATCTGACCTACTCGAGGAGGCTTACACACAGAACTCTGG 332
QY 2222 CTTGTTGTCATCTATAGGTTGAAGAGCCGATGTTCAATAGACCATCACTGAAGACC 2281
DB 331 CTTGTTGTCATCTATAGGTTGAAGAGCCGATGTTCAATAGACCATCACTGAAGACC 272
QY 2282 AAGGAGAGAACGATTCTCTCAGCAAACTTCATTTTCGAGAAAGAACTTAAGGTCGCAAG 2341
DB 271 AAGGAGAGAACGATTCTCTCAGCAAACTTCATTTTCGAGAAAGAACTTAAGGTCGCAAG 212
QY 2342 GGCTACATACAAACCGACCGTGTGTTTAAAGGAAACGAACTTGAATTAACCCAA 2401
DB 211 GGCTACATACAAACCGACCGTGTGTTTAAAGGAAACGAACTTGAATTAACCCAA 152
QY 2402 AAAAAAAAAA 2413
DB 151 AAAAAAAAAA 140

RESULT 6
US-09-270-767-1318/c
; Sequence 1318, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1318
; LENGTH: 560
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-1318

Query Match 22.9%; Score 552.4; DB 4; Length 560;
Best Local Similarity 99.8%; Pred. No. 7.5e-145;
Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1107 TCATTGATCCGTTGCGGAGCATCAGCCACCACTTGGTCTCGTCTCTTTTGTATCTGC 1166
DB 560 TCATTGATCCGTTGCGGAGCATCAGCCACCACTTGGTCTCGTCTCTTTTGTATCTGC 501
QY 1167 ACATCCCTGTTGCGCCCTTCCAGTGGGAGTGTGTACTGTCATCAAGCAGATCAACGACG 1226
DB 500 ACATCATGGTGGCCCTTCCAGTGGGAGTGTGTACTGTCATCAAGCAGATCAACGACG 441
QY 1227 AGCCGCTTTTCTGTTGCTGATCGCCATCAGTCCGGTGTACTTCGCTGGTGTATGTCG 1286
DB 440 AGCCGCTTTTCTGTTGCTGATCGCCATCAGTCCGGTGTACTTCGCTGGTGTATGTCG 381
QY 1287 GTTTGATGTTGACCCCTCAGCCCGTGTGTGTCATGCTGCGGAGTGGCTTTTCGGAC 1346
DB 380 GTTTGATGTTGACCCCTCAGCCCGTGTGTGTCATGCTGCGGAGTGGCTTTTCGGAC 321
QY 1347 TGTGGATGTTCTCTCGAAGAGATTCTGTTAAGGATGGGACAGCCATTAACGCGAG 1406

DB 320 TGTGGATGTTCTCTCGAAGAGATTCTGTTAAGGATGGGACAGCCATTAAGCGCAG 261
QY 1407 CCACCGAAGTGAAGCTGAGGATTCCATTGAGAAAGAACGCTGTACGACAAGGCTG 1466
DB 260 CCACCGAAGTGAAGCTGAGGATTCCATTGAGAAAGAACGCTGTACGACAAGGCTG 201
QY 1467 GCAAGCTGAAGCATCGTACTAAGCATGATGCCACGAGTACTGGCGTCAAGCTCCAACC 1526
DB 200 GCAAGCTGAAGCATCGTACTAAGCATGATGCCACGAGTACTGGCGTCAAGCTCCAACC 141
QY 1527 TGAAGAGTATTGTTATTTTGGCCGTTCTAATGCTGTGATGATGTTCCGCTGCCACTGCA 1586
DB 140 TGAAGAGTATTGTTATTTTGGCCGTTCTAATGCTGTGATGATGTTCCGCTGCCACTGCA 81
QY 1587 CGTGGGTGACCAAGCATGCTACTCCAGTCCCTCATTTGCTTTCACACAACTG 1646
DB 80 CGTGGGTGACCAAGCATGCTACTCCAGTCCCTCATTTGCTTTCACACAACTG 21
QY 1647 AAGATGATCCCGC 1660
DB 20 AAGATGATCCCGC 7

RESULT 7
US-09-270-767-1660/c
; Sequence 1660, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1660
; LENGTH: 560
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-1660

Query Match 22.9%; Score 552.4; DB 4; Length 560;
Best Local Similarity 99.8%; Pred. No. 7.5e-145;
Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1107 TCATTGATCCGTTGCGGAGCATCAGCCACCACTTGGTCTCTCTTTTGTATCTGC 1166
DB 560 TCATTGATCCGTTGCGGAGCATCAGCCACCACTTGGTCTCTCTTTTGTATCTGC 501
QY 1167 ACATCCCTGTTGCGCCCTTCCAGTGGGAGTGTGTACTGTCATCAAGCAGATCAACGACG 1226
DB 500 ACATCATGGTGGCCCTTCCAGTGGGAGTGTGTACTGTCATCAAGCAGATCAACGACG 441
QY 1227 AGCCGCTTTTCTGTTGCTGATCGCCATCAGTCCGGTGTACTTCGCTGGTGTATGTCG 1286
DB 440 AGCCGCTTTTCTGTTGCTGATCGCCATCAGTCCGGTGTACTTCGCTGGTGTATGTCG 381
QY 1287 GTTTGATGTTGACCCCTCAGCCCGTGTGTGTCATGCTGCGGAGTGGCTTTTCGGAC 1346
DB 380 GTTTGATGTTGACCCCTCAGCCCGTGTGTGTCATGCTGCGGAGTGGCTTTTCGGAC 321
QY 1347 TGTGGATGTTCTCTCGAAGAGATTCTGTTAAGGATGGGACAGCCATTAAGCGCAG 1406
DB 320 TGTGGATGTTCTCTCGAAGAGATTCTGTTAAGGATGGGACAGCCATTAAGCGCAG 261
QY 1407 CCACCGAAGTGAAGCTGAGGATTCCATTGAGAAAGAACGCTGTACGACAAGGCTG 1466
DB 260 CCACCGAAGTGAAGCTGAGGATTCCATTGAGAAAGAACGCTGTACGACAAGGCTG 201
QY 1467 GCAAGCTGAAGCATCGTACTAAGCATGATGCCACGAGTACTGGCGTCAAGCTCCAACC 1526
DB 200 GCAAGCTGAAGCATCGTACTAAGCATGATGCCACGAGTACTGGCGTCAAGCTCCAACC 141

QY	1527	TGAAGAGTAATTGTTATTTTGGCCCGTTCTAATGCTGTGTTGATGATGTTGCGTGTCCACTGCA	1586
Db	140	TGAAGAGTAATTGTTATTTTGGCCCGTTCTAATGCTGTGTTGATGATGTTGCGTGTCCACTGCA	81
QY	1587	CGTGGGTGACCCAGCAATGCGCTTACTCCAGTCCCTCCATTGCTTGGCTTTCACCAACAGTC	1646
Db	80	CGTGGGTGACCCAGCAATGCGCTTACTCCAGTCCCTCCATTGCTTGGCTTTCACCAACAGTC	21
QY	1647	AAGATGGATCCCGC	1650
Db	20	AAGATGGATCCCGC	7
RESULT 8			
US-09-614-221A-318			
; Sequence 318, Application US/09614221A			
; Patent No. 6723537			
; GENERAL INFORMATION:			
; APPLICANT: Karunanandaa, Balasulojini			
; APPLICANT: Yu, Jaehyuk			
; APPLICANT: Kishore, Ganesh M.			
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED			
; TITLE OF INVENTION: WITH STEROL SYNTHESIS AND METABOLISM			
; FILE REFERENCE: 16516.075			
; CURRENT APPLICATION NUMBER: US/09/614,221A			
; CURRENT FILING DATE: 2000-07-12			
; PRIOR APPLICATION NUMBER: US 60/142,981			
; PRIOR FILING DATE: 1999-07-12			
; NUMBER OF SEQ ID NOS: 626			
; SEQ ID NO 318			
; LENGTH: 2157			
; TYPE: DNA			
; ORGANISM: Saccharomyces cerevisiae			
US-09-614-221A-318			

Query Match	21.7%	Score 523.4	DB 4	Length 2157
Best Local Similarity	54.8%	Pred. No. 2.2e-136		
Matches 1212	Conservative	0	Mismatches 876	Indels 123
			Gaps	4
QY	105	TGCGCTGCTACAGCAGCCTAAATCACCCTTCGCCATCCTGCTAATCGCCTGGCTGGCGGAT	164	
DB	26	TGTCCTGTGTTTCAGACCATCTCTCAAGCTCGTCATCTTCGTGGCGATTTTGGGGCTGCCA	85	
QY	165	TTTCCTCTCGCCTCTTCGCGGCATCCGGTTTCAGTCGATTATCATGATTTGATCCGT	224	
DB	86	TATCATCAGTTTGTGTTGAGTCATCAAAATTGAGTCTATTATCCATGAATTTCGACCCCT	145	
QY	225	GGTTCAACTACCGGGCCACCGCCTACATGGTCGAGAAATGGTTGGTAGCACTTCCTCAACT	284	
DB	146	GGTTCAAATTATAGGGCTACCAATATCTCGTCAACAAATCGTTTACAGATTTTGTAACT	205	
QY	285	GGTTCCAGCAGCGCGCATCGGTATCCGCTCGGAGGATTTGGCGGCTACCGTCTATCCCG	344	
DB	206	GGTTTGACGACCGTACTCGTATCCCTCCGAGGGTTACTTGGAGGACTTTTATATCTGT	265	
QY	345	GCTGATGATTAGTTCGCGCGGAATC---CATTTGGCTGCTGCAGCTACTCAACATACGG	401	
DB	266	GTTTGAATGACGACTAGTGGTTCATCTGGCAGCCCTCGCAACTGGTTGGGCTTGGCCA	325	
QY	402	TCCATATTCGTGACATCTCGGTGTTCTTCGGCCCGGATCTTCAGTGGCTGACCTCCATCT	461	
DB	326	TTGACATCAGAAACGTTTGTGTGCTATTTCGCCACTATTTTCGGGCTCACCGCTGGG	385	
QY	462	CCACCTACTGCTGACCAAGGAGCTGTGTGTCGGGGCGCGGCTCTTCGCCCGCAGCT	521	
DB	386	CGACTTACGAATTTACGAAAGAGATTAAGATGCCAGCGCTGGGCTTTGGCTGTGGTT	445	
QY	522	TGATCGCCATCTGCTGCTGCTACATCAGTAGTGTGCTGTGCTGGCTGATACGATACGAGG	581	
DB	446	TTATAGCCATGTCGCCGGTTATATATCTAGATCAGTGGCGGGTCTTACGATATGAGG	505	
QY	582	GCATTTGCCATATTCCGCCCTGCGAGTTTCACTTCTCTGTGGGTGGCTCAGTGAGACTG	641	

D	b	506	CCATTGCCATTACACTAATTAATGTGTCACTTTTCATGTTTGGATTAAAGCCCAAAAGACTG	565
Q	y	642	GATCCGCTGTTCGTGTCGGCCGACGCGCTTGTCTACTTCTCATGCTGTCCGCGCTGGG	701
D	b	566	GCTCTATCATGCACGCAACGTGTGCAGCTTTATCTACTTCTACATGGTGTCCGCTTGGG	625
Q	y	702	GTGGCTACGTGTTTCATCATCAACCTGATACCCCTGACAGCTTCTCGTACTCTCATTTATGG	761
D	b	626	GTGGATACGTGTTTCATCACCACCTTGATCCCACTCCCATCTGTCCTTTTGTCTGATTTTGATGG	685
Q	y	762	GCAGTACTCCGCCCGCTGTGTGACCAAGCTACAGCACCTTCTCATCTCCTGGGACTGCTGT	821
D	b	686	GCAGATAATTCGTCCAACTGATCTCGCTACACCACTTGTGAGCTATTTGGAACGTGTG	745
Q	y	822	TCTCGATGCAGATCCCTTCGTGGGATTCOAACGATAGCAACAGTGAACACATGGCTG	881
D	b	746	CATCCATGCAGATCCCACTTTGTGGTTTCCTATCAAGTCTAAACGACCAATGGCCG	805
Q	y	882	CGCTGGGAGTGTGTGCTCCTTATGCGCGTGCCACCTTGTGCCCATTTTGCAGTCCGTGC	941
D	b	806	CATTGGGTGTTTTCGGTTTGATTCAGATTGTCCGCTTCGGTGACTTCGTGAAGGGCCAAA	865
Q	y	942	TGTCGGGAACAGTTCGGAAGCTGTTCACTCGCGGGAATGCTGTGGCGTTGGCG	1001
D	b	866	TCAGCAGAGTAAGTTTAAAGTCATCATGATGTTCTCTGTTTTTGTCTTGGTCCCTTG	925
Q	y	1002	TCTTTTGGCGGTGCTGCTCACCATGCTGGCGCTTGTGGCCCGTGGAGTGAAGCT	1061
D	b	926	GTGTGTCGCACTTCTGCGCTTACCTATATGGGGTTGATGGCCCTCGCATGTGTAGAT	985
Q	y	1062	TCTACTCGCTGTGGGATCTGGTACGCCAAGATCCACATTCACATTCATGTATCGGT	1121
D	b	986	TTTATTGTTATGGGATACCACTACGAAAGATCCACATTCCTATCATTTGCTCGTTT	1045
Q	y	1122	CGGAGCATCAGCCACCACTTGTCTGTTCTTTGATCTGACATCTCTGTGTGGG	1181
D	b	1046	CCGAACATCAACCCGTTTGTGGCCCGCTTCTCTTTGATACCCACTTTTGTATCTGC	1105
Q	y	1182	CTTCTCCAGTGGGAGTGTGTACTGTCATCAAGCAGATCAACACAGAGCGCTTTTCGTGG	1241
D	b	1106	TATTCCCGCGGTGATTCCTACTATTCTCGACTTGAAGACGAGACGTTTTTGCA	1165
Q	y	1242	TGCTGTACGCCATCAGTGGGTTTACTTCGCTGGTGTGATGGTGGCTTTGATGTGAACC	1301
D	b	1166	TCGCTTACTCCGTTCTGTGTTCTGTACTTTTGGCGGTGTTATGTTAGATTGATTTGACTT	1225
Q	y	1302	TCACGCCGTGTGTGCATGCTGGCCGAGTGCCTTTTTCGGGACTGTGTGATGTGTC	1361
D	b	1226	TGACACCGATCATCTGTGTGTCCGCGCGCTGCATTTGTCCAAGATATTTGACATCTACC	1285
Q	y	1362	TGCAAGAGATTCTGTCTTAAGCGAATGGGCACAGCCATAAGCCGACCCACCAAGTGGATG	1421
D	b	1286	TGGATTTCAGAACAAGTACCCGCAATACGCCATCAAACTTCGGCACTACTGG-----	1339
Q	y	1422	AAGCTGAGGATTCCATTGAGAAGAAGACGCTGTACGACAAGGCTGGCAAGCTGAAGCATC	1481
D	b	1340	-----	1339
Q	y	1482	GTACTAAGCATGATGCCACGACGAGATCTGGCGTCAGCTCCAACTGAAGAGTATTTGTTA	1541
D	b	1340	-----CCAAATTGATTGTTT	1354
Q	y	1542	TTTTGGCCGTTCTAATGCTGTTGATGATGTTTCGCTGTCCACTGCACGTGGGTGACCAACA	1601
D	b	1355	CCGGATCATTCATCTTTATTTTGATCTTTTTCGTCCTTCCATTTCTACTTTGGGTAAACAAGAA	1414
Q	y	1602	ATGCTTACTCCAGTCCCTCCATGTTCTTTGGGTTTTCCACAAACAGTCAAGATGATCCCGCA	1661
D	b	1415	CTGCATACTCTTCTCTTCTGTTGTTTGGCATCACAAACC---CCAGATGGTAAATTGG	1471
Q	y	1662	ACATTTAGACCATTTACAGAGGCTTACTACTGTGCTTTCCGAGAACACTGCCGATGATG	1721
D	b	1472	CGTTGATCGACACTTCAGGGAAGCGTACTATTGGTTAAGAAATGAATCTCATGAGACA	1531

QY 1722 CTCGCGTTATGCTCTTGGTGGATTACGGATACAGATAGCGGAATGGCAACAGCA 1781
 DB 1532 GTAAGTTGAGCGTGGTGGATTACGGTTACCAATTTGGTGGCATGGGAGACAGACCA 1591
 QY 1782 CGCTAGTGGATAATACTGTGAACAATAAGTACACATAGCGTGGTGGCAAGCAATGT 1841
 DB 1592 CTTTGTAGTGGATAACAACACGTGAACAATACTACATCGCCATCGTGGTGAAGCAATGG 1651
 QY 1842 CTTCAACCGAGGAGAGCTCTAGAAATATGACATCTCTTGACGTGGACTACGTTTGG 1901
 DB 1652 CTTCCCTTGAGAGAAATCTTACGAATTTCTAAAGAGCATGATGTGCAATTAATGCTTGG 1711
 QY 1902 TGATCTTTGGCGTGTGATCGGCTATTCTGGCGATGATATCAACAAGTTCTGTGATGG 1961
 DB 1712 TCATCTTTGGTGGTCTAATTTGGGTTTGGTGGTGAATGACATCAACAAATTTCTTGGATGA 1771
 QY 1962 TCCGAATTTCTAGGGAGAGCATCCCAAGGACATTAAGAAAGCGATTACTTTACCGACC 2021
 DB 1772 TCAGATTTAGCGAGGGAATCTGGCAGAGAGATAAAGAGCGTGNATTTCTATACCGAG 1831
 QY 2022 CGGTGAATTCAGGTAGATCCGAAGGTGCTCCGCCCTGTCTCAACTGCTTATGTACA 2081
 DB 1832 AGGAGAAATACAGATGATGCAAGGCTTCTGAGACCATGAGAACTCGCTACTTTACA 1891
 QY 2082 AATTAAGCTACTACAGATTCGGGGAATTTGAAGTTGCACTACAGAGTCCATCTGGATAG 2141
 DB 1892 AGATGCTTACAAGATTTCCCAATTAATTAATGTTGGCCAGCACTGACAGAGTGC 1951
 QY 2142 ATCCACACAGTAAAGCGCTCATTTGGGAATAAGACTTCGATCTGACCTACCTGGAGAG 2201
 DB 1952 GTCAACAAATGATCACACCATTAG-----ACGTCCACCATTAGACTACTTTCCGACGAAG 2005
 QY 2202 CCTACACACAGACACTGGCTTGTTCGCATCTATAGGTTGAGAGCGCGATGATGTTCA 2261
 DB 2006 TTTTACTTCCGAAATCGATGGTTAGAAATATCAATTAAGAGAGATGATGCCCAAG 2065
 QY 2262 ATAGACCATCACTGAAGACCAAGAGAGACGATTCCTCCAGCAACTTCA 2312
 DB 2066 GTAGAACTTTGAGGAGCGTTGGTGGTGAATTAACAGGTCTTCTACGAAACCA 2116

RESULT 9
 US-09-270-767-11648
 ; Sequence 11648, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 11648
 ; LENGTH: 1386
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-11648

Query Match 14.4%; Score 349.2; DB 4; Length 1386;
 Best Local Similarity 61.6%; Pred. No. 1.2e-87;
 Matches 558; Conservative 0; Mismatches 348; Indels 0; Gaps 0;

QY 121 CCTAATCACCTTCGCCATCTGTTAATCGCTCGCTGGCGGATTTCTCTCGCTCTT 180
 DB 467 CTTGGTGAAGTGGCAATCTCATCTCGGAGCGGTTTATCATTTGACACAGCTTGT 526
 QY 181 GCGCGTATCGTTTCGAGTCGATATTCATGAGTTTGAATCGGTGTTCACTACCGGC 240
 DB 527 CTCTGTGCTGGAATTCGAAAGCGTAATCCATGATGATTCGATCGGTACTTCACTACCGCAC 586
 QY 241 CACCGCCTACATGTTGCAGAAATGTTGGTACAACTTCTCAACTGTTTCGAGCGGC 300

DB 587 CACGCGTTTCTGGCGGAGCAGGGCTTTTACAAGTTCCACAACCTGGTTTCAATGACCCGCG 646
 QY 301 ATGGTATCCGCTCGGAGGATTTGGGCGGTACCGTCTATCCGCGCTGATGATAGTC 360
 DB 647 CTGGTATCCCTTGGGCGGATCATCGGCGGACCAATCTATCCGCGGTGATGTCACCTC 706
 QY 361 CGCGGAAATCCATTTGGCTGCTCACTGCTCAACATACCGGTCCATATTTCTGACATCTG 420
 DB 707 GGCGGCGCTGTACCGCTGATGTGCTGCTCAATGTGACCATCGACATAGGAACGTGTG 766
 QY 421 CGTGTTCCTGGGCGGATCTTCAGTGGCTGCTGACCTCCATCTCCACCTACCTGTCGACAA 480
 DB 767 CGTTCCTTGGGCGGCTTCTTCTCTGCTGACACGCTGTGACCTACGCGCTTCACAAA 826
 QY 481 GGAGCTGTGCTCGGCGGCGCGCTCTTCCGCGCCAGCTTTCATCGCCATFCGTCGCTGG 540
 DB 827 GGAGATACACAGCACTGGAGCTGGACTGCTGGCGCGCTTTGATATCCATCGTTCCCGG 886
 QY 541 CTRACATCAGTAGTCTGGTGGCTGGATCGTAGATACAGAGGCGATTTGCCATATTTGCGCCT 600
 DB 887 GTATATCTCTCGATCCGTGGGCGGATCGTAGCAATTAAGGCGATCGCCATTTTCTGCAT 946
 QY 601 GCAGTTTCACTTACTTCTCTGTGGTGGCTCACTGAGTGAAGACTGGATCCGTTTCTGCTGGC 660
 DB 947 GCTCTTCACTACTATTGTGGATCAAGCGGTAAAGACGGGACGATCTTTTGGTGGC 1006
 QY 661 CGCAGCGCTTTGTCTTACTTCTATCATGTGTCCGCTGGGCTGGCTACGTTTTCATCAT 720
 DB 1007 TATGTCGGGATTTGGCTACTTCTATATGTTCTCTGTCGGGTGGCTATGTTCTTCTGAT 1066
 QY 721 CAACCTGTATACCCCTGACGCTTCTGCTACTGCTCATTTANGGCGAGGTACTCGCGCGCTCT 780
 DB 1067 TAACCTAATCCGCTGACGCTGCTGGCGCTGATGATCACCGACGTTTCTGCGACAGAT 1126
 QY 781 GCTGACCACTACAGACCTTCTATCATCTGGAGTGTCTTCTCCATGAGATCCCTT 840
 DB 1127 CTACATAGATACAGACGCTATCTGCTGGCGACCAATCTGTCGATGAGATCTCGTT 1186
 QY 841 CGTGGGATTTCCAAACCGATACGCCAGTAAACACATGCTCGCTGGGAGTGTTTGTGCT 900
 DB 1187 TGTGGGATTTCCAAACCCATCCAGAGCTCCGAAACACATGCTGSCATGSGAACCTTTGSCCT 1246
 QY 901 CTTATGGCGGTGGCACTTGGCCATTTGCGCTGCTGCTGCTGCGCAACAGATTCG 960
 DB 1247 GTGCCAGATTCAGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1306
 QY 961 GAAGCTGTTTCACTGCTGGCGGATTTGCTGCTGGCGCTTGGCGCTTCTTGTGGCGCTGCTGCT 1020
 DB 1307 TCTGCTCTTCAAGACGTTGTTTCCAGTGTTTTCCAGTGTGTTGCTGCTGCTGCTGCTGCT 1366
 QY 1021 GCTCAC 1026
 DB 1367 GCTCAC 1372

RESULT 10
 US-09-270-767-12856
 ; Sequence 12856, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 12856
 ; LENGTH: 914
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-12856

Query Match
Best Local Similarity 11.8%; Score 284.4; DB 4; Length 914;
Matches 406; Conservative 0; Mismatches 181; Indels 3; Gaps 1;
1 1666 TTATGACGATTTTCAGAGGCTTACTAGTCTTTCAGAGAACTGCGATGATGCTCG 1725
Db 1 TTTCGATGACTTTCGCGAGGCTTACTAGTCTTTCAGAGAACTGCGATGATGCTCG 60
1726 CTTATGCTTCTGCTGGGATTTACGATACAGATAGCGGGAATGGCAACAGAACGACGCT 1785
Db 61 CATATGCTTCTGCTGGGATTTACGATACAGATAGCGGGAATGGCAACAGAACGACGAT 120
1786 AGTGATATATATAGTGGAACTAGTCAATAGTCAATAGTCAATAGTCAATAGTCAAT 1845
Db 121 AGTGATATATATAGTGGAACTAGTCAATAGTCAATAGTCAATAGTCAATAGTCAAT 180
1846 AACCGAGGAGAACTCTACGAAATATGACATCTCTTGCAGCTGGAGTACGTTTGGTGAT 1905
Db 181 TTTCGAGGAGAACTCTACGAAATATGACATCTCTTGCAGCTGGAGTACGTTTGGTGAT 240
1906 CTTTCGCGGTGATGCGCTATTTCTGGGATGATATCAACAGTTCCTGCTGGATGCTCG 1965
Db 241 TTTCGAGGAGTCACTGGCTACTCATCGGAGATATCAACAGTTCCTGCTGGATGCTCG 300
1966 AATTGCTGAGG---AGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTTTACCGACG 2022
Db 301 CATTCGCGGACACCGGATCGTGTGCGCATATCGCGAAAGGACTACTATGCGGCAA 360
2023 CGGTGATATTCAGGTAGATGCGGAGGTCTCGCGCCCTGCTCACTGCTTATGTACAA 2082
Db 361 CGGAGATTTCCGAGTGGACAGAGGGCTCACCCACACTGCTCAATTTTGTATGTACAA 420
2083 ATTAAGCTACTACAGATTCGGGGAATTGAAGTTGGACTTACAGAGGTCCATCTGGATGA 2142
Db 421 GATGTGCTACTATCTGCTTTGGGCAATGTACAGGAGGTGGCAAGGCTAGCA 480
2143 TCGCACAGTACCGGCTCATTTGGGAATAGGACTTCGATGCTGACCTACCTGGAGGCG 2202
Db 481 TCGAGTTCTGCGCGCGAGATCGGCAACAGGACTTTGAAGTGGATGCTCGAGGAGGC 540
2203 CTACACACAGAACACTGGCTTGTTCGCATCTATAGGCTGAAGAGCGCG 2252
Db 541 GTACACACGAGGACTGGCTGGTGGCATCTACAGGTTAAGGATCTGC 590

RESULT 11
US-09-513-999C-1965
; Sequence 1965, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.02.24
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1993-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1965
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 122..505
US-09-513-999C-1965

Query Match 11.6%; Score 279.6; DB 4; Length 507;

Best Local Similarity 73.1%; Pred. No. 2.2e-69;
Matches 373; Conservative 0; Mismatches 134; Indels 3; Gaps 1;
1 1606 CTATCCAGTCCCTCCATTCCTTCCTTCCTCCCAACAGTCAAGTGGATCCCGCAACAT 1665
Db 1 CTATCCAGTCCCTCCATTCCTTCCTTCCTCCCAACAGTCAAGTGGATCCCGCAACAT 57
1666 TTATGACGATTTTCAGAGGCTTACTAGTCTTTCAGAGAACTGCGATGATGCTCG 1725
Db 58 CTTATGCTTCTGCTGGGATTTACGATACAGATAGCGGGAATGGCAACAGAACGACG 117
1726 CTTATGCTTCTGCTGGGATTTACGATACAGATAGCGGGAATGGCAACAGAACGACGCT 1785
Db 118 AGTATGCTTCTGCTGGGATTTACGATACAGATAGCGGGAATGGCAACAGAACGACT 177
1786 AGTGATATATATAGTGGAACTAGTCAATAGTCAATAGTCAATAGTCAATAGTCAAT 1845
Db 178 GGTGATATATATAGTGGAACTAGTCAATAGTCAATAGTCAATAGTCAATAGTCAAT 237
1846 AACCGAGGAGAACTCTACGAAATATGACATCTCTTGCAGCTGGAGTACGTTTGGTGAT 1905
Db 238 TAATGAAACACAGGCTTATAAATCATGAGGACTCTAGATGTAGATTATGTTTGGTAT 237
1906 CTTTCGCGGTGATGCGCTATTTCTGGGATGATATCAACAGTTCCTGCTGGATGCTCG 1965
Db 238 TTTCGAGGAGTTCCTGCTGATGATATCAACAGTTCCTGCTGGATGCTCG 357
1966 AATTGCTGAGGAGGATCCCAAGGACATTAAGGAAAGCGATTACTTTTACCGACGCG 2025
Db 358 GATAGCTGAGGAGGACATCCCAAGGACTTCGGAAGTGAATTTTACCCCAAGG 417
2026 TGAATTCAGGATAGATGCGGAGGTGCTCGCGCCCTGCTCAACTGCTTATGTACAAAT 2085
Db 418 AGAATTCGCTGATAGCAAGGAGGATCCCTACTCTTCTGGAATGCTTATGTATAAAT 477
2086 AAGTACTACAGATTCGCGGGAATTGAAGTT 2115
Db 478 GTCATACATGAGATTTGGAGAAATGAGCT 507

RESULT 12
US-09-248-796A-3089
; Sequence 3089, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1999-02-13
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3089
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (25)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unk
US-09-248-796A-3089

Query Match 10.1%; Score 244.4; DB 4; Length 867;
Best Local Similarity 60.1%; Pred. No. 2.2e-58;
Matches 445; Conservative 0; Mismatches 286; Indels 9; Gaps 2;
1 1534 TATTGTTATTTTGGCCCTTCTAATGCTGTTGATGATGCTGCTGCTGCTGCTGCTGCT 1593
Db 48 TTTCGTTTACTGACATTTACATTTTACCTTTTCTACTTTTCTACTTTTCTACTTTTCTACT 107

QY 1594 GACCAGCAATGCCCTACTCCAGTCCCTCCATTGCTTGGCTTCCCAACAGTCAAGATGG 1653
DB 108 AACATCGATGCTTATTCATCACCACAGTGTGTTAGCATCCAGAAC--CCAGATGG 164
QY 1654 ATCCGCAACATTTAGACGATTCAGAGAGGTTACTACTGCTTTTCAGAAACCTGC 1713
DB 165 CTCACACATATCATTCGATGATATAGAGAGGCTTATCTGTTAAGAAATGAATACACC 224
QY 1714 CGATGATGCTCGGCTTATGCTTGTGGGATACAGATACAGATAGCGGGAATGGCAA 1773
DB 225 AGAAGATGCCAAGTTAAGGCTTGTGGGATATGTTATCAATCGGGGATAGGCTGA 284
QY 1774 CAGAACGACGCTAGTGGATTAATATAGTGGAAATATGACATCTTTGACGTGGAATA 1893
DB 285 TAGAACACACATTTGATTAACAATACATGGAATAACACACATATTTGCCACTGTTGGTAA 344
QY 1834 GGCAATGCTTCAACCGAGGAGAGTCTCTACGAAATATGACATCTTTGACGTGGAATA 1893
DB 345 GGCATGCTTCCCTCGAATGTGCTGATGAATTTTGACACACAGATGTTGATTA 404
QY 1894 CGTTTGGTGTATCTTTGGCGGTGTGATCGGCTATTTCTGGGATGATATCAACAGTTCCT 1953
DB 405 TGTCTTATGATATTTTGGAGGTTTATTTGGGTTATTTCTGCTGATGATATTAACAAATTTCT 464
QY 1954 GTGATGCTCGAATTTCTGAGGAGAGATCCCAAGGACATTAAGGAAAGGATTAATT 2013
DB 465 ATGATGTTGAATTTCTGAGGTATCTGGCTGTATGAATCAAGAAAGAGACTACTT 524
QY 2014 TACCGACGGGTGAATTCAGGTAGATGCCAGGTGTCTCGGCGCTCTCAATGCTCT 2073
DB 525 TACTGACCGAGGAGATATAAGTGGGATAAAGATGATCACTGTCGCAATGAAGAATTTCTTT 584
QY 2074 TATGTACAAATTAAGCTACTACAGATTCGGGGAATTAAGTTGGACTACAGAGGTCATC 2133
DB 585 GATGTATAAGTTATCTGATCATAGATTCAGTGAATTTGAGGTAGAGATGGTGTGA 644
QY 2134 TGGATATGATCGCACGTAACCGGTGATGCGGATGATGCGGCTCTGATCTGACCTACCT 2193
DB 645 TAGAGTTAGAAACCAACAAATCCAGCCCAATGAAGTACCGAAATGAATGT-----TGT 698
QY 2194 GGAGAGGCTTACACACAGACAACTGGCTTGTTCGCTCTATAGGTTGAAGAAGCGCA 2253
DB 699 TGAAGAGCCTTACATCAGAAATTTGGATTTGAGAAATTTACAAATTAAGATTTGA 758
QY 2254 TGAGTTCAATAGACCATCAC 2273
DB 759 TAATGTTGGTAGATTTAC 778

RESULT 13

US-09-328-111-332

; Sequence 332, Application US/09328111

; Patent No. 6262333

; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.

; APPLICANT: Steinmann, Kathleen E.

; APPLICANT: Astle, Jon H.

; APPLICANT: Burgess, Christopher C.

; APPLICANT: Bushnell, Steven E.

; APPLICANT: Carroll III, Eddie

; APPLICANT: Catino, Theodore J.

; APPLICANT: Derti, Adnan

; APPLICANT: Ford, Donna M.

; APPLICANT: Lewis, Marcia E.

; APPLICANT: Monahan, John E.

; APPLICANT: Schlegel, Robert

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; FILE REFERENCE: CDD-257 (US)

; CURRENT APPLICATION NUMBER: US/09/328,111

; CURRENT FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: US 60/088,801

; EARLIER FILING DATE: 1998-06-10

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 332

; LENGTH: 616

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(616)

; OTHER INFORMATION: n = A,T,C or G

US-09-328-111-332

Query Match 7.2%; Score 173; DB 3; Length 616;

Best Local Similarity 64.0%; Pred. No. 1.9e-38;

Matches 260; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 533 GTGCTGGCTACATCAGTAGTGGTGGCTGGATCGTACGATAACGAGGCAATTCGCATA 592

DB 5 GTACCAGGCTACATATCTCGGTCACTAGCTGATCCCTTTGATAATGAAGCATTTGCTATT 64

QY 593 TTGCGCCCTCAGTTTCACTTCTCTGTTGGTGGCTCAGTGAAGACTGATCCGTTTC 652

DB 65 TTGCACTTCACTTACATATCTTATGGTAAATCTGTAAATCTGGGTGAGTTT 124

QY 653 TGTGCGGCGCAGCGCTTTGCTTCTTCTACATGTTGCTCGGCTGGGTGGCTAGCTG 712

DB 125 TGGACAATGTGCTGCTTATCTTCTTCTATGTTCTCTGCTGGGTGGTTATGTA 184

QY 713 TTCTCATCAACTGATACCTCTGCACTCTTCTGTTCTGTTCTGTTCTGTTCTGTTCTG 772

DB 185 TTCTCATCAACTTATTTCTCACTGCAATGTTTGTGTTGTTGTTGTTGTTGTTGTTG 244

QY 773 CGCGCTCTCTGACAGCTTACAGACCTTCTACATCTCTGAGCTGCTGTTCTCCATGCG 832

DB 245 AAGAGAGTCTACATAGCATATAGCACTTTCTACATGTTGGTTTAAATATATCAATGCG 304

QY 933 ATCCCTTCTGTTGGATTCACCGATAGCAGCAGTGAACACATGCTGGCTGGGAGTG 892

DB 305 ATACCTTTTGTGGGATTCGACCAATCAGAACAGTGAACATGTCAGCTTTCAGGTGC 364

QY 893 TTGTGCTCTTATGCGGCTGGCCACCTTGGCCATTTTGCAGTCCG 938

DB 365 TTGCACTCTGCAAGCTTAACTTTCTTTCAGTATCTGAGAACCG 410

RESULT 14

US-09-270-767-27262

; Sequence 27262, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7328-054

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 27262

; LENGTH: 503

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

US-09-270-767-27262

Query Match 7.1%; Score 171.2; DB 4; Length 503;

Best Local Similarity 59.4%; Pred. No. 5.5e-38;

Matches 290; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 539 GGCTACATCAGTAGTGGTGGCTGGATCGTACGATACGAGGCAATTCGCATATTCGCC 598

DB 2 GGGTATATCTTCTGATCCGTCGGGATCGTACGCAATGAGCAATGAGCATTCGCATTTTCGC 61

QY 599 CTGCACTTCACTTCTCTGTTGGTGGCTGCTGAGTGAAGACTGATCCGTTCTGTTCTGTCG 658

```
Db 62 ATGCTCTTCACTACTATTGTGGATCAAGCGGTAAAGACGGGACAGATCTTTGGTCG 121
QY 659 GCGGAGCGGCTTTGTCTTACTTCTACATGTGTCCGCTGGGTGGCTAGCTTTCATC 718
Db 122 GGTATGTGGGCAATGGGCTACTTATATGTCTCTCTCGTGGGTGGCTATGTCTTCCG 181
QY 719 ATCAACCTGATACCCCTGCACTCTTCGTACTGTCTCATATTGGGCACTACTCGCCCGT 778
Db 182 ATTAACCTAATCCGCTGCAGTGTGGCGTGTATGATCAACCGAAGCTTCTCGCACAG 241
QY 779 CTGCTGACAGCTACAGACCTTCTACATCTGGAGTGTCTTCTCATGACAGATCCC 838
Db 242 ATCTACATAGCATACAGACCGTATCTACTGCGTGGCACTTCTGTGATGCAAGATCTG 301
QY 839 TTGCTGGGATTCACACCCATAGCACCACTAGTGAACACATGGCTGGGAGTGTGTG 898
Db 302 TTGTGGGATTCACACCCATCCAGAGCTCCGAACACATGTGGCACTGGAACTTTGGC 361
QY 899 CTCCTTATGGCGTGGGCACTTGGCCATTTGCAGTCCGTGTGTGGCGCAAGATTC 958
Db 362 CTGTGCCAGATTCACGCTTTTGTGACTATCTGGCTGGCACTTCCCAAGGATCACTT 421
QY 959 CGGAAGCTTTTCATCGTGGCGGATTTGTGTGGCGTGGCTGTGTGGCGGCTGTG 1018
Db 422 GATCTGCTCTCAAGACGTTGGTTCCAGTGTTTTGACTGTGGTTCGTCGGGTACC 481
QY 1019 GTGCTCAC 1026
Db 482 CTGCTCAC 489
```

RESULT 15

```
US-09-385-982-213
; Sequence 213, Application US/09385982
; Patent No. 6282334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCNDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: PastSeq for Windows Version 3.0
; SEQ ID NO 213
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-385-982-213
```

```
Query Match 6.2%; Score 150.2; DB 3; Length 487;
Best Local Similarity 57.3%; Pred. No. 4.2e-32;
Matches 272; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 797 ACCTTCTACATCTCTGGGACTGTCTTCTCCATGACAGATCCCTTCGTGGGATTCACACCG 856
Db 1 ACTGTTTACTGCTGGGCACTATCTTCTATGACAGATCTCTTTGTGGGTTTCCAGCCT 60
QY 857 ATACGCACAGTGAACACATGGCTGGGTGGTGTGTCTTATGGCGGTGGCC 916
Db 61 GTCTTTTATACAGACATGGCAGCCTTTGGGTCCTTTGTCTCTGCCAGATCCATGCC 120
QY 917 ACCTTGGCCATTTTCAGTCCGCTGTGGCGCAACGAGTTCCGGAAGCTGTTCATGTC 976
Db 121 TTGTGGATTACCTCGGAGCAAGTTGAATCCACAAATTTGAAGTCTTTTCCGGAGC 180
```

```
QY 977 GCGGATTTGCTGTGGGCTTGGCGTCTTTGTGGCGGCTGTGTGTGCTCACCATGCTGGGC 1036
Db 181 GTCATCTCTCTGGTAGGCTTTGTCTCTTCTCACCGTGGAGCTCTCTCATGCTGACAGGA 240
QY 1037 GTTGGGCCCCGTGGAGTGGACGTTTCTACTGTGTGGGATACCTGGCTACGCCAAGATC 1096
Db 241 AAAATATCTCCCTGGACGGGCGTTTCTACTCTCACTGTGGATCCCTCTTATGCTAAGAAC 300
QY 1097 CACATTCCCATCATTTGTCATCGTGTGGAGCATCAGCCCAACCACTTGGTTCTGTTCTTC 1156
Db 301 AACATCCCATCATTTGTTCTGTGTCTGAGCATCAGCCCAACCACTTGGTTCTTACTAT 360
QY 1157 TTTGATCTGACATCCCTGGTGTGGCCCTTCCAGTGGAGTGTGTGTCTGCTCATCAGCAG 1216
Db 361 TTGACCTGACAGCTCCCTGCTTTCATGTTTCCAGTTGGCCTCTATTACTGCTTTAGCAAC 420
QY 1217 ATCAACGACGAGCGGCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1271
Db 421 CTGTCTGATGCCCGGATTTTATCATGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 475
```

Search completed: December 15, 2004, 11:16:18
Job time : 196.945 secs

Blank
page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 07:01:53 ; Search time 1208.01 seconds
(without alignments)
11031.546 Million cell updates/sec

Title: US-10-028-384-7

Perfect score: 2417

Sequence: 1 tctagcgaagaagtgtcg.....ccaaaaaataaaaaaa 2417

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/PCTUS_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2417	100.0	2417	15	US-10-028-384-7
2	960.6	39.7	2710	15	US-10-028-384-3
3	955.8	39.5	2481	15	US-10-028-384-1
4	636	26.3	3046	18	US-10-417-375-95
5	635.4	26.3	2232	15	US-10-128-714-7139
6	632.8	26.2	2472	15	US-10-171-581-112
7	632.8	26.2	2472	15	US-10-028-384-11
8	632.8	26.2	2472	15	US-10-172-118-742
9	632.8	26.2	2472	15	US-10-342-887-742
10	631.2	26.1	3093	18	US-10-417-375-92
11	631.2	26.1	3094	15	US-10-028-384-3
12	628.6	26.0	5404	18	US-10-417-375-99

13	628.6	25.0	5827	18	US-10-417-375-97	Sequence 97, Appl
14	620.2	25.7	2466	16	US-10-320-797-2305	Sequence 2305, Ap
15	619	25.6	2839	18	US-10-425-115-150745	Sequence 150745,
16	615	25.4	2681	16	US-10-425-114-14408	Sequence 14408, A
17	612.4	25.3	2779	17	US-10-437-963-99904	Sequence 99904, A
18	562	23.3	1848	15	US-10-128-714-2139	Sequence 2139, Ap
19	523.4	21.7	2157	9	US-09-801-368-387	Sequence 387, App
20	523.4	21.7	2157	18	US-10-793-639-318	Sequence 318, App
21	521.8	21.6	2733	15	US-10-028-384-5	Sequence 5, Appl
22	519.8	21.5	2256	15	US-10-032-585-6323	Sequence 6323, Ap
23	497	20.6	1969	15	US-10-128-714-1139	Sequence 1139, Ap
24	497	20.6	2603	15	US-10-128-714-6139	Sequence 6139, Ap
25	497	20.6	3969	15	US-10-128-714-1339	Sequence 1339, Ap
26	497	20.6	4603	15	US-10-128-714-5139	Sequence 5139, Ap
27	467	19.3	1209	15	US-10-106-698-330	Sequence 330, App
28	467	19.3	1209	16	US-10-264-237-412	Sequence 412, App
29	466.4	19.3	1543	10	US-09-574-879-133	Sequence 133, App
30	466.4	19.3	1543	10	US-09-305-736-133	Sequence 133, App
31	466.4	19.3	1543	10	US-09-818-683-133	Sequence 133, App
32	466.4	19.3	1543	11	US-09-818-683-133	Sequence 133, App
33	466.4	19.3	1543	16	US-10-621-401-133	Sequence 133, App
34	430.2	17.8	1828	10	US-09-945-527-62	Sequence 62, Appl
35	403	16.7	2660	16	US-10-264-049-630	Sequence 630, App
36	382.6	15.8	3141	18	US-10-425-115-130787	Sequence 130787,
37	373.8	15.5	2882	16	US-10-320-797-1305	Sequence 1305, Ap
38	373.8	15.5	4738	16	US-10-320-797-305	Sequence 305, App
39	357.4	14.8	3197	16	US-10-424-599-111541	Sequence 111541,
40	311.6	12.9	1728	16	US-10-424-599-122476	Sequence 122476,
41	310.8	12.9	1114	16	US-10-296-115-629	Sequence 629, App
42	291.2	12.0	2244	17	US-10-437-963-48342	Sequence 48342, A
43	279.8	11.6	887	16	US-10-424-599-77697	Sequence 77697, A
44	247.4	10.2	430	9	US-09-878-178-717	Sequence 717, App
45	247.4	10.2	430	13	US-10-046-935-717	Sequence 717, App

ALIGNMENTS

RESULT 1
US-10-028-384-7
; Sequence 7, Application US/10028384
; Publication NO. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAULT, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 2417
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AF132552
; DATABASE ENTRY DATE: 1999-04-27
; RELEVANT RESIDUES: (1)..(2417)
US-10-028-384-7

Query Match 100.0%; Score 2417; DB 15; Length 2417;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAGCGAAGAAGTGTGCTTGCATTTCAGATCGGTATAATTTTCAGTTACTGCTG 60
Db 1 TCTAGCGAAGAAGTGTGCTTGCATTTCAGATCGGTATAATTTTCAGTTACTGCTG 60
QY 61 GAATTTGGACATGAATCGGACGCCGAGATGCTGAACAGCAAGTGGCTGCTACAGCAG 120
Db 61 GAATTTGGACATGAATCGGACGCCGAGATGCTGAACAGCAAGTGGCTGCTACAGCAG 120

QY 121 CCTAATCACTTCGCCATCTCTGCTAAATCGCTGGCTGGCCGGATTTTCTCTCGCTCTT 180
DB 121 CCTAATCACTTCGCCATCTCTGCTAAATCGCTGGCTGGCCGGATTTTCTCTCGCTCTT 180
QY 181 CGCGTCACTCGTTTCGAGTCGATATCCCATGAGTTTATCCGTTGGTTCAACTACCGGC 240
DB 181 CGCGTCACTCGTTTCGAGTCGATATCCCATGAGTTTATCCGTTGGTTCAACTACCGGC 240
QY 241 CACCGCTACATGCTGAGATGCTTGGTACACTTCTCACTGCTTGAAGGGCC 300
DB 241 CACCGCTACATGCTGAGATGCTTGGTACACTTCTCACTGCTTGAAGGGCC 300
QY 301 ATGGTATCCGCTCGCGAGATTTGGGGGGTACCGTCTATCCCGGCTGATGATACGTC 360
DB 301 ATGGTATCCGCTCGCGAGATTTGGGGGGTACCGTCTATCCCGGCTGATGATACGTC 360
QY 361 CGCGGATCACTTGGCTGCTGACGCTACTCAACATACCGTCTCATATTCGTCGATCTG 420
DB 361 CGCGGATCACTTGGCTGCTGACGCTACTCAACATACCGTCTCATATTCGTCGATCTG 420
QY 421 CGTGTCTCGCGCGGATCTTTCAGTGGCTGACCTCACTCCACCTACTCTGACCAA 480
DB 421 CGTGTCTCGCGCGGATCTTTCAGTGGCTGACCTCACTCCACCTACTCTGACCAA 480
QY 481 GGAGTGTGGTCCGGGGCGCGCTTTCGCGCGAGCTTCACTGCCATGCTGCTGG 540
DB 481 GGAGTGTGGTCCGGGGCGCGCTTTCGCGCGAGCTTCACTGCCATGCTGCTGG 540
QY 541 CTACATCAGTAGTGGTGGTGGTACGATACGAGGGATTTGCCATATTCGCCCT 600
DB 541 CTACATCAGTAGTGGTGGTGGTACGATACGAGGGATTTGCCATATTCGCCCT 600
QY 601 CGAGTTCACCTACTTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 660
DB 601 CGAGTTCACCTACTTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 660
QY 661 CGAGTTCACCTACTTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 720
DB 661 CGAGTTCACCTACTTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 720
QY 721 CAACTGATACCTGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 721 CAACTGATACCTGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 781 GCTGACCACTGACGACCTTCTACATCTGCGGACTGCTGCTTCCATGACGATCCCTT 840
DB 781 GCTGACCACTGACGACCTTCTACATCTGCGGACTGCTGCTTCCATGACGATCCCTT 840
QY 841 CGTGGATTCACCGGATCCGACGACGACGACGACGACGACGACGACGACGACGACGAC 900
DB 841 CGTGGATTCACCGGATCCGACGACGACGACGACGACGACGACGACGACGACGACGAC 900
QY 901 CTTATGCGCTGGCGACCTTGGCGGATTTGCGAGTCCGCTGCTGCTGCGCAAGAGTTCG 960
DB 901 CTTATGCGCTGGCGACCTTGGCGGATTTGCGAGTCCGCTGCTGCTGCGCAAGAGTTCG 960
QY 961 GAAGTGTTCATGCTGGCGGATTTGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1020
DB 961 GAAGTGTTCATGCTGGCGGATTTGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1020
QY 1021 GCTCACCATGCTGGCGGCTTGGCGGCTGGGAGTGGAGCGCTTCTACTCGCTGCGGATAC 1080
DB 1021 GCTCACCATGCTGGCGGCTTGGCGGCTGGGAGTGGAGCGCTTCTACTCGCTGCGGATAC 1080
QY 1081 TGGCTACCGCAAGATCCACATTCGATCCGCTGGAGCATCCGCTGGAGCATCCGCTGG 1140
DB 1081 TGGCTACCGCAAGATCCACATTCGATCCGCTGGAGCATCCGCTGGAGCATCCGCTGG 1140
QY 1141 TTGGTTCCTGCTTCTTGTGATCGACATCTGCTGGTGGCGCTTCCGAGTGGGAGTGG 1200
DB 1141 TTGGTTCCTGCTTCTTGTGATCGACATCTGCTGGTGGCGCTTCCGAGTGGGAGTGG 1200

QY 1201 GTACTGCATCAAGCAGATCAACGAGCGCGTTCCTGCTGGTGTAGCCCATCAGTGC 1260
DB 1201 GTACTGCATCAAGCAGATCAACGAGCGCGTTCCTGCTGGTGTAGCCCATCAGTGC 1260
QY 1261 GGTACTTCTCGCTGGTGTAGTGTGATGTGACCTTCAAGCTGAGTGTGATGTGAT 1320
DB 1261 GGTACTTCTCGCTGGTGTAGTGTGATGTGACCTTCAAGCTGAGTGTGATGTGAT 1320
QY 1321 GCTGGCGGAGTGGCTTTTGGGACTGTTGGAGTGTTCCTGCAAGAGGATTCGTCTAA 1380
DB 1321 GCTGGCGGAGTGGCTTTTGGGACTGTTGGAGTGTTCCTGCAAGAGGATTCGTCTAA 1380
QY 1381 GCGAATGGGACAGCCATTAAGCCGACCAAGCTGAGTGTGATGAGTGTGATGATGAT 1440
DB 1381 GCGAATGGGACAGCCATTAAGCCGACCAAGCTGAGTGTGATGAGTGTGATGATGAT 1440
QY 1441 GAGAGAGCGCTGACGACAGCGTGGCAAGCTGAGCATCTGCTACTAGCATGATGCCA 1500
DB 1441 GAGAGAGCGCTGACGACAGCGTGGCAAGCTGAGCATCTGCTACTAGCATGATGCCA 1500
QY 1501 GCAGGATCTGCGCTGAGCTTCAACCTGAGAGTATTTTATTTTGGCGCTTCTAATGCT 1560
DB 1501 GCAGGATCTGCGCTGAGCTTCAACCTGAGAGTATTTTATTTTGGCGCTTCTAATGCT 1560
QY 1561 GTTGTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
DB 1561 GTTGTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
QY 1621 CATTTGCTTGGCTTCCACACAGTCAAGATGGATPCCCGCAACATTTTACAGATTTTCA 1680
DB 1621 CATTTGCTTGGCTTCCACACAGTCAAGATGGATPCCCGCAACATTTTACAGATTTTCA 1680
QY 1681 AGAGGCTTACTGCTGCTTCCAGACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
DB 1681 AGAGGCTTACTGCTGCTTCCAGACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
QY 1741 GATTTACGATACAGATAGCGGAAATGCAACAGACGCTAGTGGATTAATAATAC 1800
DB 1741 GATTTACGATACAGATAGCGGAAATGCAACAGACGCTAGTGGATTAATAATAC 1800
QY 1801 GTGGAACATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
DB 1801 GTGGAACATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
QY 1861 CTACGAAATATGACATCTTTCAGCTGAGTACGCTTGTGCTGCTGCTGCTGCTGCTGCTG 1920
DB 1861 CTACGAAATATGACATCTTTCAGCTGAGTACGCTTGTGCTGCTGCTGCTGCTGCTGCTG 1920
QY 1921 CGGCTATTCTGCGGATGATCAACAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
DB 1921 CGGCTATTCTGCGGATGATCAACAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
QY 1981 GCATCCCAAGGACATTAAGAAAGGATTTACCTTACCGACCGCGTGAATTCAGGGTGA 2040
DB 1981 GCATCCCAAGGACATTAAGAAAGGATTTACCTTACCGACCGCGTGAATTCAGGGTGA 2040
QY 2041 TGGCGAGGCTGCTCGGCGCTGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
DB 2041 TGGCGAGGCTGCTCGGCGCTGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
QY 2101 CGGGAATTTGAAGTGGACTACAGAGTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
DB 2101 CGGGAATTTGAAGTGGACTACAGAGTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
QY 2161 CATTTGGGAATAGGACTTCCATCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
DB 2161 CATTTGGGAATAGGACTTCCATCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
QY 2221 GCTTGTTCGATCTATPAGGCTGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
DB 2221 GCTTGTTCGATCTATPAGGCTGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
QY 2281 CAAGGAGAGAACGATTTCTCTCCAGCAAACTTCTTTCGAGAGAAAGAACTCTAAGCTGCTGCA 2340

Db 2281 CAGGAGAGAAAGATTCTCCAGCAAACTTATTCGAGAAAGAACTCTAAGCGTCGCA 2340
Qy 2341 GGCTACATACGAAACCGACCGGTTGTTGTTAAAGGAAAGAAAGAAACCTTGAATAAAGCCCA 2400
Db 2341 GGCTACATACGAAACCGACCGGTTGTTGTTAAAGGAAAGAAAGAAACCTTGAATAAAGCCCA 2400
Qy 2401 AAAAAAARAAAAAARAAA 2417
Db 2401 AAAAAAARAAAAAARAAA 2417

RESULT 2

US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)...(2469)

US-10-028-384-3

Query Match 39.7%; Score 960.6; DB 15; Length 2710;
Best Local Similarity 64.8%; Pred. No. 7, 4e-285;
Matches 1495; Conservative 0; Mismatches 794; Indels 18; Gaps 4;
Qy 102 AGGTGGTGGCTACAGCAGCCTAATCAGCTGGCCACTCTGCTAATCGCTGGCTGGCTGGCCG 161
Db 258 AGCGGCGGGGGGAGAGTGGTGGTCTCTCTTCCACCATCTCTTCTGGCTGGCTGGCTGGCCG 317
Qy 162 GATTTTCCTCTCGCTCTCTCGCGCTCATCCGCTTCGAGTGCATATCATGAGTTTGCATC 221
Db 318 GCTTCAGCTCGCGCTCTTCGCGCTCATCCGCTTCGAGGAGCATCATCCAGGTTCCACC 377
Qy 222 CGTGGTTCACTACCGGCGCACGGCTCATGCTGCGAGAAATGGTTGGTACAACTTCTCA 281
Db 378 CGTGGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGATTCTATGAGTTCTAA 437
Qy 282 ACTGGTTCCAGCAGCGCGCATGGTATCCGCTCGGCGGAGTTCGCGGGGTACCGTCTATC 341
Db 438 ATTGGTTTATGAAAGAGCATGTATCCCACTGGGAGAAATAGTGGTGGCAGCGTTTACC 497
Qy 342 CCGGCTGTATGATCTCGCGGAAATCCATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 401
Db 498 CAGGGTTGATGATAACAGCTGGCCCTTATTCATTGGATTAAATACATTGAAATACAG 557
Qy 402 TCATATTCGTCGACATCTCGTGTTCCTCGGCGCCGATCTTCAGTGGGCTGACCTCCATCT 461
Db 558 TTCACATAAGATGTGTGTATTCCTTTGACCAAACTTTTAGCGGCTTTACATCCATAT 617
Qy 462 CACCTTACTGCTGACCAAGAGCTGTGGTCCGCGGGCGCGGCTCTTTCGCGCCAGCT 521
Db 618 CTACGTTCTGTCTAATAGAGAACTGTGGAACCAAGGAGCAGGACTTCTAGCTGCTGCT 677
Qy 522 TCATCGCCATCGTGGCTGCTACATCAGTAGTGGTGGCTGGATCGTACGATAACGAGG 581
Db 678 TCATTGCTATCGTACCAGGGTACATATCTCGGTGAGTGGCGGATCTTGTATGATGAAG 737
Qy 582 GCATTGCCATATTCGCCCTGAGTTCCACTTACTTCTGCGGTGGCTCAGTGAAGACTG 641

Db 738 GCATTGCCATTTTTCGCGCTTCAGTTCACTTACTTATCGGTAAGACTCTGTGAAGACCG 797
Qy 642 GATCCGCTGTCTGTGGTGGCGCCAGCGCTTGTCTTACTTCTTACTTCTTACTTCTTCTTCT 701
Db 798 GGTCTGTCTTCTGGACAAATGTGCTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCT 857
Qy 702 GTGGCTACGTGTCTTATCATCACTGATACCCCTGACGCTTCTGCTGCTGCTGCTGCTGCT 761
Db 858 GAGGTATGTGTCTTATCATCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 917
Qy 762 GCAGGTACTCGCGCGCTGCTGCTGACCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 821
Db 918 AGAGGTACAGCAAGAGAGTCTTACATAGATATAGCACTTGTGTATCTTGTGGTTTAAAT 977
Qy 822 TCTCCATGAGATCCCTTCTGCTGGATTCNAACGATACGACCACTGAGTGAACATGCTG 881
Db 978 TATCCATGAGATACCTTCTTGTGGATTCAGGCAATCAGAAAGAGGAGCATGGCAG 1037
Qy 882 CGCTGGGAGTGTGTGCTCTTATGCGCGTGGCCACTTGGCCACTTTCAGTCCGCTGC 941
Db 1038 CTGCAAGGTCTTTCGCGCTGCTGCAAGCTTACGCTTTCGAGTATCTGAGAGACCGGT 1097
Qy 942 TGTCCGCAACGAGTTCGGAAGCTGTCTCATCTGCGCGGATTCGCTGCTGCTGCTGCTG 1001
Db 1098 TGACAAAACAGGAGTTCAGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1157
Qy 1002 TCTTTTGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1061
Db 1158 TGTTCCTTAGTGTCTATCTTCTGACATACACAGTTTATTTGACCATGAGTGGCAGGT 1217
Qy 1062 TCTACTGCTGCTGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1121
Db 1218 TTTTACTATGAGGATCTGGGTATGCAAAATACACATTCCAATTTTGTGATGCTAGTGT 1277
Qy 1122 CGGAGCATCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1181
Db 1278 CTGAACATCAGCTTACGACATGGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1337
Qy 1182 CTTTCCAGTGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1241
Db 1338 CTTTCCAGAGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1397
Qy 1242 TCTCTACGCCATCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1301
Db 1398 CTCTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1457
Qy 1302 TCAGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1361
Db 1458 TGACCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1517
Qy 1362 TGCAAGAGATTCGCTTAAAGCAATGGGCAAGCCTAAGCGAGCGCCAGCGAGTGGATG 1421
Db 1518 TGGGGGATGA-----CATGAAAGGGGAAACCCACCTCTGGAG-----GACAGCAGTG 1565
Qy 1422 AGCTGAGGATTCATTGAGAAAGAGCGCTGTACGACAGGCTGGCAGCTGAGCATC 1481
Db 1566 ATGAGGATGACAAAGAAACCCAGGAACTTGTATGACAGGCGAGGTAAGTGAAGGAGC 1625
Qy 1482 GTACTAAGCATGATGCCAGCAGGATCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1541
Db 1626 ATGTGACAGCAAGAGAAACCTTGAAGAGGCTTGGGCGCCCAACATCAAAAGCATTTGA 1685
Qy 1542 TTTTGGCGGTCTTAACTGCTGTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1601
Db 1686 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1745
Qy 1602 ATGCTTACTCCAGCCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1661
Db 1746 ACGCTTACTCCAGTCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1802
Qy 1662 ACATTTTGAAGATTCAGAGAGCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1721

Db 1803 ATATATTAGATGATTTTAGAGACCGTACTTTTGGCTGAGACAAAACACGGATGACACG 1862
 Qy 1722 CTGCGTTATCTCTGTGGGATTTACGGATACAGATAGGGGAATGGGAAACAGAGCA 1781
 Db 1863 CCCGGTCACTCTGTGGGACTACGGCTATCAGATTTGTGGCATGGGCAACAGAGCA 1922
 Qy 1782 CGCTAGTGGATAAATACGTGGAAACAATAGTCACATAGCGCTGGTGGCAAGGAATGT 1841
 Db 1923 CTCTGTGGATTAACAACCTGGAAACAACAGCACATCGACTGGTGGAAAGCTATGT 1982
 Qy 1842 CTTCAACGAGGAGAGTCTTACGAATATGACATCTTTGAGTGGACTACGTTTGG 1901
 Db 1983 CTTCAATGAACCGCGCTTATAAATCATGAGTTCCTTGTGATGTCGATTAATGTGTGG 2042
 Qy 1902 TGATCTTTGGCGGTGATCGGCTATTCTGGCGATGATATCAACAAGTTTCTTGGATGG 1961
 Db 2043 TTAATTTTCGGAGAGTGNATGGCTATTCGGGGACGATATCAACAAGTTTCTTGGATGG 2102
 Qy 1962 TCCGAATTCGTAGGGAGAGATCCCAAGGACATTAAGGAAGCGATTAATTTACCGACC 2021
 Db 2103 TCAGGATAGCTGAAGGGGAGCATCCCAAGAGACATCCGGGAAGGTGACTATTTTACCCGAGC 2162
 Qy 2022 GCGGTGAATTCAGGCTAGATCCCAAGGTGCTCCGGCCCTGCTCAACTGCTTATGTACA 2081
 Db 2163 AGGAGAGTTCGGAGTAGACAAGCTGGGTCTCTTACTCTGTAACTGCTTATGTATA 2222
 Qy 2082 AATTAAGCTACTACAGATTCGGGAATGAAGTTGACTACAGAGTCCATCTGGATATG 2141
 Db 2223 AAATGTCTACTACAGATTTGGAGAAATGCAGCTAGATTTTCGCACTCCCGGAGCTTTG 2282
 Qy 2142 ATCGCACACGTAACCGCGTCAATTCGGAAATGAAGACTTCGATCTGACCTACCTGGAGAGG 2201
 Db 2283 ACCGACACGTAATGCTGAGATTGGAAATGAAGCATTAATTCAGCATTTTGGAGGAG 2342
 Qy 2202 CTTACACACAGACACTGGCTTTGCGCATCTATAGGGTGAAGAGCCGATGAGTTCA 2261
 Db 2343 CTTTATACATCAGACACTGGCTTTGCTAGGATATATAAAGTGAAGACACCTGACACAGGG 2402
 Qy 2262 ATAGACCATCTGTAAGACCAAGAGAGACGATTCCTCCAGCAAACTTCATTTGAGAA 2321
 Db 2403 AGACACTAGTTCACAAACCTCGATCCACACATCGTCCCAACAGAGATTTGTCAA 2462
 Qy 2322 AGAAGCTTAAGCGTGGCAAG---GGCTACATACGAACAGACCGGTTGTTGTAAGGAA 2378
 Db 2463 AGAAGACTACTAAAGGAGCGTGGCTACGTTAAATAAATAGCTAGTGTGTTAAAGAGGCA 2522
 Qy 2379 AACGAACTTGAATTAACCCAAAAA 2405
 Db 2523 AGAAGACCTTAAGAGACTGTTTAA 2549

RESULT 3
 US-10-028-384-1
 ; Sequence 1, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREULT, Claude
 ; APPLICANT: MCBRIDE, Kevin
 ; TITLE OF INVENTION: Mammalian SMP, gene sequence and uses thereof in cancer therapy
 ; FILE REFERENCE: 5600-74
 ; CURRENT APPLICATION NUMBER: US/10/028,384
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 2481
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2481)
 ; OTHER INFORMATION:

US-10-028-384-1
 Query Match 39.5%; Score 955.8; DB 15; Length 2481;
 Best Local Similarity 64.8%; Pred. No. 2.1e-283;
 Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;
 Qy 102 AGTGCTGGCTACAGAGCCCTAAATCAGCTTCGGCATCTCTTAATCGCTGGCTGGCCG 161
 Db 191 AGCCGGCTGGGTGGGACGCTTCTCTCTCATTACCACTCTCTCTCTGGCTGGCTGGCCG 250
 Qy 162 GATTTTCCTCTCTGGCTCTCTGGCGCTATCGCTTCGAGTGCATTAATCATGAGTTGATC 221
 Db 251 GCTTCAGCTCGGCGCTCTCTGGCGCTATCGCTTCGAAAGCATCATCCAGAGTTGAC 310
 Qy 222 CGTGGTTCAATACACGGGCCACCGCTACATGCTGTCAGAAATGGTTGTCACAACTTCTCA 281
 Db 311 CGTGGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGGTCTCTAGATTTTAA 370
 Qy 282 ACTGGTTTCAGAGCGCGCATGGTATCGCTCGGAGAGATTGTGGGGTACCGTCTATC 341
 Db 371 ATTGGTTTGTAAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGGTACTGTATTACC 430
 Qy 342 CCGGCTGATGATTAAGTTCGGCGGAATCCATTTGGCTCTGCAGCTACTCAACATACCGG 401
 Db 431 CAGGTTGATGATAACCGCTGGCTTATTCATTGGAATTTAAATACATTGAACATACTG 490
 Qy 402 TCATATTCGTGACATCTGCGTTCTGGCGCGCATCTTCAGTGGCTGACCTCCATCT 461
 Db 491 TCCACATAAGAGACGTATGTGTGTTCTTTCGACCAACTTTTAGCGGCTTACATCTATAT 550
 Qy 462 CACCTTACCTGCTGACCAAGAGAGCTGTGCTCGGCGCGCGGCTCTTCGCGCGCAGCT 521
 Db 551 CTACTTCTCTTACAGAGAGACTTTGGAACCAAGAGAGAGAGCTTTTAGCTGCTTGT 610
 Qy 522 TCATCCCATCTGCTGGCTACATCAGTAGTGGTGGCTGGATGATGATGATGATGATGATG 581
 Db 611 TTATTGCTATTGTACAGGCTACATATCTCGGTGAGTAGTGGATCTTTGATAATGAAG 670
 Qy 582 GATTCCTCATATTCGCTGCTGAGTTCACCTACTTCTGTTGGTGGCTGCTGATGAAAGCTG 641
 Db 671 GCATTGCTATTTTGCACCTTCATATCTACTATTTATGGTAAATCTGTAAAGACTG 730
 Qy 642 GATCGTGTCTGCTGGCGCGCGAGCGCTTTGCTCTACTTCTACATGGTTCGCGCTGGG 701
 Db 731 GGTCACTTTTGGGCAATGTGCTGCTTATCTATTTCTATATGGTCTCTGCTTGGG 790
 Qy 702 GTGGCTACGTGTTCAATCACTGATACCCCTGCAGCTCTTCGTACTGCTCATTTATGG 761
 Db 791 GTGGTTATGTTTATCATCAATCTTATTCACATGATGTTTGTGTTACTGATGC 850
 Qy 762 GAGGTACTCGCGGCTGCTGACAGCTACAGACCTTCTACATCTGGGACTGCTGT 821
 Db 851 AGAGATACAGCAAAAGAGCTTACATAGCATATAGCACTTTCTACATTGGGTTTAAATAT 910
 Qy 822 TCTCCATGAGATCCCGCTTCTGGGATTCGAACCGATACGACCAAGTGAACACATGGCTG 881
 Db 911 TATCAATGAGATACCTTTTGGGATTCAGCAATTCAGAAACAAGTGAACACATGSCAG 970
 Qy 882 CGCTGGAGTGTGTGCTCTTATGGCGGTGGCACCTTGGCCATTTGACGTCCTGTCG 941
 Db 971 CTGAGGTGCTTGTGATGCTGCAAGCTTATGCTTTCTTCAGTATCTGAGAGACCGAT 1030
 Qy 942 TGTGCGCAACAGATTCCGGAAGCTGTTTCATCGTGGCGGATTCCTGTTGGCGTTGGCG 1001
 Db 1031 TAAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGTATCACTAGCTGACGTTGCTG 1090
 Qy 1002 TCTTTGTCGCTGCTGGTGTCTCACAATGCTGGGCTGTCGCGGCTGGAGTGGAGCTG 1061
 Db 1091 TGTCTCTAGTGTCTATTTGACTTATACAGTTACATTGCAACATGGAGTGGAGCT 1150
 Qy 1062 TCTACTCGCTGTGGATATGCTACGCAAGATCCCAATTTCCCATCTATTGCAATCCGTGT 1121
 Db 1151 TTTATTCAITGGGATGCTGGTATGCAAAATACACATTTCCAAATATTGTCATCAGTGT 1210

1122 CGGAGCATCAGCCACCACTGGTTCTCGTCTCTTTGATCTGCACATCCTCGTGTGG 1181
1211 CTGAGCATCAACCTACGACTGGGTGCTCTCTCTTTGATCTACATATCTTTGATGTA 1270
1182 CTTCCAGTGGAGTGTGGTACTGATCAAGAGAGATCAACGAGCGGCTTTCTGTGG 1241
1271 CTTCCAGCGAGCGCTTTGGTCTGATCAAAATATCAAGATGAAGAGATTTGTTG 1330
1242 TGCTGTAGCCATCAGTCCGGTTTACTTCGCTGTGTGTGATGGTCTTTGATGTGACCC 1301
1331 CTCTATATGCATCACTGCTCTCTCTTTGCTGAGTGTGTTGGTCTGATGTGACTT 1390
1302 TCACCCGGTGTGTGATGCTGTGCGGAGTGGCTTTTGGGACTGTTGGATGTGTTC 1361
1391 TGACTCCAGTGTGTGTGATGCTGTGCAATTTGCTTTTCAAAATGTTTGTGAGCACTATT 1450
1362 TGCAAGAGATTCGCTCTAGCGAATGGGCACAGCCATAAGCGCAGCCACCGAAGTGGATG 1421
1451 TGGGGATGA-----CATGAAAGGGAATCCACTGTGGAG-----GACAGAGTG 1498
1422 AAGCTGAGATTCATTTGAGAGAGAGCTGTGTAGCAAGAGCTGGCAAGCTGAAGATC 1481
1499 ATGAGGATGACAAAAGAACCAAGGAAATTTGTATGATAAGCGCAGGTAAAGTGAGGAAAC 1558
1482 GTACTAAGCATGATCCCGCAGGAGATCTGGCGTCAGCTCAACCTGAAGAGTATTGTTA 1541
1559 ATGCACTGACAGGAAAGAACTGAGAGGATTTAGGCCCTAATATATAAAGCATTTGTCA 1618
1542 TTTTGGCGTCTTAAGCTGTGTGATGATTTGCTGCTCCACTGCACTGCGGTGACGAGCA 1601
1619 CCATGTTGATGCTGATGATTTGATGATGTTTGTGTCCACTGTGCTGCTGCTGCTGCTG 1678
1602 ATGCTCTACTCAGTCCCTCCATTTGCTTGGCTTTTCCACAGCTCAAGATGGATCCGCA 1661
1679 ATGCTCTACTAGTCCAGTGTAGTCTGGCTCATACAA---TCATGATGGCACCAGGA 1735
1662 ACATTTAGAGATTTTACAGAGGCTTACTACTGGCTTTCCGAGAACACTGCGGATGATG 1721
1736 ATATCTTAGATGATTTTACAGAGCTTACTTTTGGCTAAGCGCAAAATACAGATGAACATG 1795
1722 CTCGGGTATGCTTTGGTGGATTTACGATACAGATAGCGGAATGGCAAAACAGAACGA 1781
1796 CAGAGATGATGCTTTGGTGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGACTA 1855
1782 CGCTAGTGATTAATACGTGGAAACAATAGTCAATAGCTGCTGGTGTGGCAAGCAATGT 1841
1856 CGTTGGTGGATTAATAACACCTGGATTAACAGCCACATAGCACTGGTGGGAAAGCTATGT 1915
1842 CTTCAACGAGGAGAGCTCTACGAATTTATGACATCTCTTGAGCTGCACTAGCTTTTG 1901
1916 CTTCTAATGAACAGCAGCCCTATAAATCATGAGGACTCTAGATGTAGATTTGTTTGG 1975
1902 TGATCTTTGGCGGTGTGATCGGCTATTTCTGCGATGATATCAACAAAGTTCCTGTGATGG 1961
1976 TTATTTTGGAGGGGTATTTGGCTATTTCTGTGTATGATATCAACAAATTTCTCTGGATG 2035
1962 TCCGAATTCCTGAGGAGAGATCCCAAGGACATTAAGGAAGCGATTTACTTTACCGACC 2021
2036 TTAGATAGCTGAAGAGAAACATCCCAAGACATTCGGAAGATGACTATTTTACCCAC 2095
2022 GGGGTGAATTCAGGGTAGATCGGAGGTGTCCGGCCCTGCTCAACTGCTTTATGTACA 2081
2096 AGGGAGAAATCCGTGTAGCAAAAGCAGGATCCCTCTACTTTGTTGAATTTGCTTTATGATA 2155
2082 AATTAAGCTACTACAGATTTCCGGGAATTTGAATTTGGACTACAGAGGTCCATCTGGATATG 2141
2156 AATGTCTACTACAGATTTGGAGAAATGCACTGGATTTTGTACACCCCGAGTTTGG 2215
2142 ATCGCAACGTAACCGCTCAATTTGGGAATAGGACTTCGATCTGACTGACCTCGAGAGG 2201
2216 ACCGAACGTAATGCTGAGATTTGGAATAGGACATTAATTAACAAATTTGGAAGAG 2275

2202 CCTACACACAGAACACTGGCTTGTTCGCTATCTATAGGTGAAGAAGCCGATGAGTTCA 2261
2276 CTTTACATCAGAACACTGGCTTGTAGGATATATAAGTAAAGCACCTGATAACAGGG 2335
2262 ATAGACCATCTCTGAAGCAAGAGAGAGATTCCTCCAGCAAACTTCTCATTTTCAGAA 2321
2336 AGACATTTAGATCAAAACCTCGAGTCCACCAACATTTTCCCAAAACAGAAAGTATTTCTCA 2395
2322 AGAATCTCTAAGCGTCCGAAG---GGCTACATACGAAACCGACCGTGTGTTGTTAAGGAA 2378
2396 AGAAGACTACAAAGGAGCGTGGCTACATTAATAAATAGCTGGTGTTTTAAAGGAA 2455
2379 AACGAACCTTGAATAAAC 2397
2456 AGAAATATCTAAGAAGAC 2474

RESULT 4
US-10-417-375-95
; Sequence 95, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95
; LENGTH: 3046
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-95

Query Match 26.3%; Score 636; DB 18; Length 3046;
Best Local Similarity 58.3%; Pred. No. 1.1e-184;
Matches 1249; Conservative 0; Mismatches 800; Indels 93; Gaps 4;

118 CAGCCTAATCACTCCCTCCCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 177
180 CACATTTTGAAGCTTCTCATCTCTGCAATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 239
178 CTTCCCGCTCATCCGTTTCAGTGCATATCCAGATGATTCAGATGATTCAGTGCATGCTGCT 237
240 GTTTGCTGCTGCTGAGATTTGAAAGTGTATCCATGAGTTGATCCGCTACTTTAAATATCG 299
238 GCCACCGCCTACATGCTGAGATGGTGGTACAACTTCTCAACTGCTGCTGCTGCTGCTGCTGCT 297
300 GACTACCAAGTTCTGCTGAGGAGGGTTTATTAATTTCCATTAATGCTGCTGCTGCTGCTGCTGCT 359
298 GCAATGCTATCCGCTCGGCAAGATTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 357
360 AGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419
358 GTCCGGCGGGAATCCATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 417
420 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479
418 CTGCGTGTCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 477
480 CTGCT 539
478 CAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537
540 CAAGAGCTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 599
538 TGGCTACATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
600 TGGATATATCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659
598 CTTGAGTTCACCTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 657

Query Match		26.3%;	Score 635.4;	DB 15;	Length 2232;
Best Local Similarity		58.9%;	Pred. No. 1.4e-184;	Indels 75;	Gaps 6;
Matches 1248;		Conservative 0;	Mismatches 196;		
QY	137	ATCCGTGTAATGCGCTGCGCGGATTTTCCTCTCGCCCTCTTCGCGCTATCCGCTTC	196		
DB	76	ATCCGTGTAATGCGCTGCGCGGATTTTCCTCTCGCCCTCTTCGCGCTATCCGCTTC	135		
QY	197	GAGTGTATATCCAGTATGATCCGCTGCTTCACTACCGGCGCACCGCTACATGCTG	256		
DB	136	GAGTGTATATCCAGTATGATCCGCTGCTTCACTACCGGCGCACCGCTACATGCTG	195		
QY	257	CAGATGTGTGATCAACTTCTCAACTGTTCAGCAGCGCGCATGGTATCCGCTCGCG	316		
DB	196	CAGATGTGTGATCAACTTCTCAACTGTTCAGCAGCGCGCATGGTATCCGCTCGCG	255		
QY	317	AGGATGTGTGCGGCTACCGTCTATCCCGGCGCTGATGATGCTCCGCGCAATCCATGG	376		
DB	256	CGTGTACCGGCTGCGGCTATATCCCGGCTCTATGCTGAGCGCGGCTGATCTACCAT	315		
QY	377	CTGCTGACGTAATCAATACCGGCTCAATTCGTGACATCTGCTGCTGCTGCTGCTG	436		
DB	316	ATCTTGGCATTCCTTACTATCCCGGCTGATATTCGCAACATCTGCTGCTGCTGCTG	375		
QY	437	ATCTTCACTGCGCTGACCTCATCTCCACTACCTGCTGACCAAGGAGCTGCTCGCG	496		
DB	376	GGATTTCTCGGCTGACCTCATCTCCACTACCTGCTGACCAAGGAGCTGCTCGCG	435		
QY	497	GGC---GGCGGCTCTTCGCGCGGCTTCACTCCCATTCGCTGCTGCTGCTGCTGCTG	553		
DB	436	CCATCTGCGAGTCTTCTTGCAGCAGCTTTCATGGGAATCGCCCTGCTGCTGCTGCTG	495		
QY	554	TCGGTGGCTGATCGTACGATAACGAGGCGATTCGCATATTCGCCCTGCGAGTTCACCTAC	613		
DB	496	TCAGTGTCTGAGCTGATCAATCAACGAGGATTCGCATATTCGCCCTGCGAGTTCACCTAC	555		
QY	614	TTCTGTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	673		
DB	556	TTCTGTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	615		
QY	674	TCCTACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	733		
DB	616	TTCTAGCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	675		
QY	734	CTGCAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	793		
DB	676	CTGCAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	735		
QY	794	AGCAGCTTCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	853		
DB	736	ACCATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	795		
QY	854	CGGATACGCAAGTGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	913		
DB	796	CCTATCCGAAACAGCGACCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	855		
QY	914	GCACCTTGGCCATTTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	973		
DB	856	GCCTTGGCCGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	915		
QY	974	GTGCGCGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1033		
DB	916	GCATGATCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	975		
QY	1034	GGCGTGTGGCCCTGGAGTGGAGCTTCTACTCGCTGCTGCTGCTGCTGCTGCTGCTG	1093		
DB	976	GGAGTGTATGCTTCTGAGCGGCGGATTTCTACTCTTGTGCTGCTGCTGCTGCTGCTG	1035		
QY	1094	ATCCAGTTCCTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1153		
DB	1036	ATCCAGTTCCTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1095		
QY	1154	TTCTTTGATCTGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1213		

DB	1096	TTCTTCGATCTGAACTTCTTGATCTGGCTTTTCCCGCAGGTGCTTACATGCTTCGGT	1155		
QY	1214	CAGATCAACGACGAGCGGTTTTCGTGGTCTGTAGCCCATAGTGGGTTTACTTTCGCT	1273		
DB	1156	GACCTCAAGGACGAGCATGCTTTCGTCAATATCTACTCGGTCTTCCGAGCTACTTTC	1215		
QY	1274	GGTGTGATGGTGGCTTTGATTTGACCTCAGCCCGGTGGTGGCATGCTGGCGGAGTG	1333		
DB	1216	GGTGTATGTTCCGACTAATGCTGACCTTGACCCCTATTGTGTGTGTTGCCGCTGCTG	1275		
QY	1334	GCCTTTTCGGGACTGTGTGGATGTTCCTCTCAAGAGGATTCGTCTAAGCGAATGGGACA	1393		
DB	1276	CGCTGTGCTCCATTTCTCGACACCTATATG-----	1305		
QY	1394	GCCATAAGCGCAGCAGCAGGATGAGCTGAGGATTCATTCAGAGAAGAGCGCTG	1453		
DB	1306	GCGACTACCTCCCGACACCGCGTCTGAAGC-----	1338		
QY	1454	TACGACAAAGGCTGCAAGCTGAGAGCATCTGTAAGCATGATCCCGCAGAGGATATGGC	1513		
DB	1398	AAAAAGAAATGAAGACTCGTCTTCAACACCTCTTCGCTCAAGTTCGGAAGCCCAATGTGA	1398		
QY	1514	GTCAAGCTCCAAACCTG---AAGAGTATTTGTTATTTTGGCCGTTCTAATGCTGTGATG	1570		
DB	1399	ATCACTTCCCATGTTTCTAAGATTATAGTCAGCGGCTCTGTTGCTCTACCTGCTCTG	1458		
QY	1571	TTCCGTGCTCACTGACGCTGGGTGACGCAANTGCTTCTCACTGCTGCTGCTGCTGCTG	1630		
DB	1459	TTTGTGGGCACTGCACTCGGTTTACATCGAATGCACTACTTCTCTCTCCGTTGCTTG	1518		
QY	1631	GCTTTCACAAACAGTGAAGTGGATCCCGCAACATTTAGACGATTTACAGAGGCTTAC	1690		
DB	1519	GCTAGTCCGATGCTCT---GACGGAAGCCAAATACATCATTCACGATTTCTGAGGCTTAC	1575		
QY	1691	TACTGCTTTTCGAGAACACTCGCGATGATGCTCGGTATGCTGCTGCTGCTGCTGCTG	1750		
DB	1576	TACTGCTTCTCGAATACTCTCTCAGACGCCAAATCATGTCTGCTGCTGCTGCTGCTG	1635		
QY	1751	TACAGATAGCGGGAATGGCAACAGAACGACGCTAGTGATTAATACTGCGAACAT	1810		
DB	1636	TATCAATCGTGCATGCGGACCGCCCAACCTTGTGTCACAAACACCTCGAACAC	1695		
QY	1811	AGTCAATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1870		
DB	1696	ACCCATATTTGCTACGCTTGGTAAAGCGATGAGCTCAGCGAGGAAGTCACTACCCCATC	1755		
QY	1871	ATGACATCTCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1930		
DB	1756	CTCGCGCAGCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1815		
QY	1931	GGCGATGATCAACAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1990		
DB	1816	GGCGATGATCAACAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1875		
QY	1991	GACATTAAGGAAGCGATTTACTTTACCGACCGCGGTGAATTCAGGGTAGATGCGAAGGT	2050		
DB	1876	GAGGTAAAGAGCGGACTTCTTTACTGACCGCGGTGAATATCGTGTGCGAGTGGAGCG	1935		
QY	2051	GCTCGGCGCTGCTCAACTGCTTATGTACAATTAAGCTACTACAGATTCGGGGAAATTG	2110		
DB	1936	ACCCCAACTATGCGCAACAGCTTGTGTATATAAATGCTTTATTACA-----ATTTC	1986		
QY	2111	AAGTGGGACTACAGAGGTCTCATCTGATATGATCGCACAGTAAACCGCGCTCATTTGGGAAT	2170		
DB	1987	ACTCTCTCTTCCGCTCGGCGCAAGCTGCGACCGCGTCCGTGGGTCAAAACTTCCACACA	2046		
QY	2171	AAGGACTTTCGATCTGACCTACTGAGAGGCGCTTACACCAAGAACTCTGCTGCTGCTG	2230		
DB	2047	GAGGCGCTCTGCTCTTACACTCGAAGAGCTTTTCGAGCGAGAACTGGATCATTCGT	2106		
QY	2231	ATCTATAGGTTGAAGAGC 2249			

Db 2107 ATCTACAAGGTCAAGGATC 2125

RESULT 6

```

US-10-171-581-112
; Sequence 112, Application US/10171581
; Publication No. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mac Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10/171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 112
; LENGTH: 2742
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: L38961
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-112

```

Query Match	26.2%;	Score 632.8;	DB 15;	Length 2472;
Best Local Similarity	58.2%;	Pred. No. 9.4e-184;		
Matches 1247;	Conservative	0;	Mismatches 802;	Indels 93;
Gaps	4			
QY	118	CAGCCATAACCTCCGCACTCTGCTGAATCGCTGGCTGGCCGGATTTCTCTCGCT	177	
Db	151	CACACTTTGAAGCTTCTCATCTGTCAATGGCTGCTGATATATCCTTCTCACCTGCT	210	
QY	178	CTTCGCCGTATCCGTTTCGAGTCGATATCCATGAGTTTGATCCGTGGTTCAACTACCG	237	
Db	211	GTTCCTCTCTCGAATTTGAAGCTGTATCCATGAGTTTGATCCGTACTTTAATATTCG	270	
QY	238	GGCCACCCGCTACATGGTGCAGAAATGGTGGTCAACTCTCCTCAACTGGTTCGACGAGCG	297	
Db	271	GACTACCAAGGTTCTGGCTGAGGAGGGGTTTATAAATCCATACTGTTTGATGACCG	330	
QY	298	CGCATGGTATCGCTCGCAGGATGTGGCGGTACCGTCTATCCCGGCTCGATGATAC	357	
Db	331	AGCCTGGTACCCCTTTGGGACGAATCATTTGGAGGAACAATTTACCCAGGTTTAATGATCAC	390	
QY	358	GTCCGGCGGAATCCAAATGGCTGTCTGCACGTACTCAACATACCGGTCCATTCGTGACAT	417	
Db	391	CTCTGCTGCAATCTACCAATGTACTCCATTTTTCACATCACCATCGCAATTCGGAATGT	450	
QY	418	CTGGGTGTTCTGGGGGGGATCTTCAGTGGCGCTGACCTCGATCTCAACTCTCGCTGAC	477	
Db	451	CTGTGTGTCTCGGCCCTCTCTCTCCCTCTTCACTCCATCGTCACTACCTCTCTTAC	510	
QY	478	CAAGAGAGCTGTGTCCGCGCGCGCGCCCTCTTTCGGCGCGAGCTTCATCGCCATCGTGCC	537	
Db	511	CAAGAGCTCAAGGATCGAGGGCTGGGCTCTTGTCTGTGCCATGATTGCTGTAGTTCC	570	
QY	538	TGGCTACATCAGTAGGTGCGTGGCTGATCGTAGATACGAGGGCATTCGCATATTCGC	597	
Db	571	TGGATATATCTCCCGATCTGTGGCTGGCTCCTATGATAATGAAGGAATGCCATCTTTG	630	
QY	598	CTCGAGTTCACCTACTTCTGTGGGTGCGCTCAGTGAAGATGGATCCGTGTTCTGTGTC	657	
Db	631	CATGTACTACCTACTACATGTGGATCAAGGCAGTAAGAGCTGGTTCCATCTGTGGGC	690	
QY	658	GGCCGACCGCTTTGTCCTACTTCTACATGGTGTCCGCGTGGGTGGCTAGCTGTTCAAT	717	
Db	691	AGCTAAGTGTGCCCTTGCTTATTTCTACATGGTCTGTCATGGGAGGTTATGTGTTCTCT	750	
QY	718	CATCAACTGATACCCCTGCACGCTCTTCGTACTGCTCATTTATGGGCAGGTACTCCCGCG	777	

Db	751	GATCAACTTAATTCCTCTCCAGCTCTCTCGTGTGATGTCACAGGCCGTTTCTCTCAACG	810
QY	778	TCGTCTACCAAGCTTACAGCACTTCTACATCTCTGGGACTGTGTTTCTCCATGACAGATCC	837
Db	811	GATCTATGTGGCCTACTGTACTGTTTACTTGCCTGGTACTATCTTTCTAGGCAGATCTC	870
QY	838	CTTGTGGGATTCACACCGATACGACAGTGAACACATGCTGCTGCGTGGAGTGTGTTGT	897
Db	871	CTTTGTGGTTCACGCTGCTCTTTCATCAGAGCACATGCGAGGTTTGGGGCTTTGG	930
QY	898	GCTCTTATGCGCGTGCGCCACCTTGCGCAATTGTCAGTCCGTGTGTGTCGCGCAACGAGTT	957
Db	931	TCTCTGCCAGATCATGSCCTTTGTGGATTACCTGCGCAGCAAGTTGAATCCACAACAATT	990
QY	958	CGGAAAGCTTTTCATCGTGGCGGATTTGCTGTGGCGCTTGGCGTCTTTGTGGCGCTCGT	1017
Db	991	TGAAGTTCTTTTCGGAGCGTCATCTCTGTGAGCTTTGTCCCTTCTACCGTGGGAGC	1050
QY	1018	GGTGCTCACCATGCTGGGCGTTGTGGCCCGCTGGAGTGGACGCTTCTACTCGCTGTGGGA	1077
Db	1051	TCTCTCATGCTGACAGGAAAAATATCTCCCTGGACGGGGCTTCTACTCACTGCTGGA	1110
QY	1078	TACTGGCTAGCCAAAGATCCACATTCCTCCATCATTCGATCCGCTCGGAGATCAGCCCA	1137
Db	1111	TCCCTTTATGCTAAGAACAAACATCCCAATCAATGCTTCTGTGCTGAGCATCAGCCAC	1170
QY	1138	CACTTGHTTCTCGTTCCTTTTGATCTGCATCTCTGTGTGCGCTTCCCACTGGGAGT	1197
Db	1171	AACCTGTGCTCTCATCTATTTTGACCTGCAGCTCCTCGTCTTCATGTTCCAGTTGCGCT	1230
QY	1198	GTGTACTGTCATCAAGCAGATCAAGCAGAGCGCGTTTTCTGTGTGCTGTACGCCATCAG	1257
Db	1231	CTATTACTGCTTTAGCAA CCGTGTGATGCCCGAATTTTATCAATGATGTGTGTGAC	1290
QY	1258	TGCGGTTTACTCGCTGTGTGATGTGGTGTGATGTGACCTCACGCGGTGFGTG	1317
Db	1291	CAGCATGTACTTTTTCAGCTGTAATGTGGCTCTAATGCTAGTGTGGCACTGTATTAG	1350
QY	1318	CATCTGCGCGAGTGCGCTTTTCGGGACGTGTGATGTGTTCTGCAAGAGGATTCGTC	1377
Db	1351	CATTCTCTGGCAATTGGAGTCTCCAGGTGCTG-----	1384
QY	1378	TACGGAATGGGCACACCCATAAGCGACGCCACGGAAGTGATGAAGCTGAGGATTCAT	1437
Db	1385	-----TCCAC	1389
QY	1438	TGAGAAGAAGACGTGTAGCACAAGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGC	1497
Db	1390	ATACATGAGAACTTGGACATATAGTGCGCCACACAAAGAGCAGAGCAACAGGATTC	1449
QY	1498	CCAGCAGATCTAGGCGTCAGCTCCAACTGAAAGATATCTTTATTTTGGCGCTTCTAAT	1557
Db	1450	CACCTACCCCTATTAAAGATT-----GAAGTGCAAGTGGGATGATCTGGTCACTGGCTTT	1503
QY	1558	GCTGTTCATGATGTTGCTGTCCACTGCACTGGGTGACCAAGCAATGCCCTACTCCAGTCC	1617
Db	1504	CTTTCTCATCACCTACACCTTTTCATTCAACTGTGGTGACCAAGTGGGCTACTCTTCTCC	1563
QY	1618	CTCAATTGCTTGGCTTTCCAAACAGTCAAGATGGATCCCGCAACATTTTACAGATTT	1677
Db	1564	GTCAATTGTACTACTGTGCC-----GTGGTGGGATGGCAGTAGGATCATATTTGATGACIT	1620
QY	1678	CAGAGAGGCTTACTACTGGGCTTTTCGCAAACTCCCGATGATGCTCGCGTTATGTCTTG	1737
Db	1621	CCGAGAAGCATATTATTGGCTTCGTCATPAACTCCAGAGATCGGAAGTCAATGCTCG	1680
QY	1738	GTGGGATACGGATACAGATAGCGGAATGGCAAAACAGAACGACGCTAGTGGATATAA	1797
Db	1691	GTGGGATTTAGGCTATCAGATTTACAGTTACAGCTATGCGAAACCGAAACAATTTTAGTGACATAA	1740
QY	1798	TACGTGAAACAATAGTACATAGCGCTGGTTGGCAAGCAATGTCTTCAACCGAGAGAA	1857
Db	1741	CACATGAAATAATCCCAATTTCTCGAGTAGGGGAGGCAATGGCTCCACAGAGAAAA	1800

1858 GTCCACGAAATATGACATCTCTGACGTGAGTACGTTTGGTGTATCTTGGCGGTG 1917
1801 AGCCTATGAGATCATGAGGAGCTCGATGTAGCTATGTGTGTCTATTTTGGAGCCT 1860
1918 GATCGGCTATCTGCGGATGATATCAACAAGTTCTGTGGATGTCGGAATTCGTGAGG 1977
1861 CACTGGGTATCTCTGATGATATCAACAAGTTCTTGGATGTCGGAATTCGAGGAG 1920
1978 ---AGAGCATCCCAAGGACATTAAGGAAAGGATTAATTTACCGACCGGTGAATTCAG 2034
1921 CACAGATACAGCAACATATCAGAGAAATGATTAATTAATCTCACTGGGAGTTCCG 1980
2035 GGTAGATCCGAAAGTGTCCGCGCTCTCAACTGCGCTTAATGATCAAAATTAAGTACTA 2094
1981 TGTGAGCCGTGAAGTTCTCCAGTGTCTCAAGTGTCTCAAGTGTCTCAAGATGTCTACTA 2040
2095 CAGATTCGGGGAAATGAAGTTGAGTACAGAGGTCCTATCGATATGATCGCACGATAA 2154
2041 TGGCTTTGACAGGTTTACAGAGCCAGGTCCTCCAGGCTTGACCGTGTCCGAAA 2100
2155 CGCGCTANTGGAAATAGGACTTCGATCTGACCTTACCTGAGGAGGCTTACACACAGA 2214
2101 TGTGAGATTTGGAAATAAAGACTTTGAGCTTGATGTCCTGGAGGAAGCTATACACAGA 2160
2215 ACACTGGCTTGTTCGATCTATAGGTTGAAGGTCGAGGCGCATCA 2256
2161 ACATTTGGCTGTGTCAGATATACAGGTAAGGACTTGATTA 2202

RESULT 7

US-10-028-384-11
; Sequence 11, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002219
; DATABASE ENTRY DATE: 2000-10-31
; RELEVANT RESIDUES: (1)..(2472)

US-10-028-384-11

Query Match 26.2%; Score 632.8; DB 15; Length 2472;
Best Local Similarity 58.2%; Pred. No. 9.4e-184;
Matches 1247; Conservative 0; Mismatches 802; Indels 93; Gaps 4;

118 CAGCCTAATACACCTTCGCGATCCTCTAATTCGCTGGCTGGCGCGATTTCTCTCGCCT 177
151 CACACTTTTGAAGCTTCTCAATCTGTCAATGGCTGCTGATATATCCCTTCCTCCACTCGTCT 210
178 CTTGCGCGTATCCCGTTTCGAGTCGATTTATCCATGAGTTTGCATCCGTTCAACTACCG 237
211 GTTTGCTGCTCGAGATTTGAAGTTTATCCATGAGTTTGCATCCGTTCAATATATCG 270
238 GGCCACCGCTACATGGTGCAGATGGTGTGATCAACTCTCACTGCTGTCAGCAGG 297
271 GACTACAGGTTCTTGGCTGAGGAGGGTTTATATAATTTCCATAACTGGTTTATGACCG 330
298 CGCATGGTATCCGCTCGGAGGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTAC 357
331 AGCCTGGTACCCTTTGGGACGAATCATTTGGAGGAACAATTTACCCAGGTTTATGATCAC 390

358 GTCCGGCGGAATCCATTTGGCTGTCTGACGTACTCAACATACCGGTCCATATTCGTGACAT 417
391 CTCTCTCTCAATCTACCAATGACTTCCATTTTCCACATCACCATCGACATTCGGATGT 450
418 CTGCGTGTCTCTGGCGGAGTCTTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGAC 477
451 CTGTGTGTCTCTGGCGGCT 510
478 CAGGAGAGTGTGTCTCCGCGCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 537
511 CAAAGAGCTCAAGATGACAGGGCTGGGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 570
538 TGGCTACATCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 597
571 TGGATATATCTCCCGATCTGTGGCTGGCTCTATATGATAATGAAGGATTCGCACTCTTTG 630
598 COTGAGATTCACCTACT 657
631 CATGCTACTCACCCTACT 690
658 GCGCGAGCGCTTGTCT 717
691 AGCTAAGTGTGCGCTTGTCT 750
718 CATCAACCTGATACCCCTGACAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 777
751 GATCACTTAATCT 810
778 TCTGCTGACAGCTACAGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 837
811 GATCTATGTGGCTACTGTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 870
838 CTTCTGGGATTCGAAACCGATACGACAGTGAACATCGCTGCGCTGGGAGTGTGTGT 897
871 CTTTGTGGTTCAGGCTGCT 930
898 GCTCTTATGCGCTGGCGCTTGTGGCTTGTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 957
931 TCTCTGCGAGATCCATGCTTGTGGATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 990
958 CCGGAGCTGTTCT 1017
991 TGAAGTCTTCT 1050
1018 GGTCTCACCATGTGGCGTGTGGCGCTGTGGCGCTGTGGCGCTGTGGCGCTGTGGCGCTGT 1077
1051 TCT 1110
1078 TACTGGCTACGCAAGATCCACATTCCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1137
1111 TCCCTCTTATCT 1170
1138 CACTTGGTCT 1197
1171 AACCTGGTCT 1230
1198 GTGGTCT 1257
1231 CTTTCT 1290
1258 TGGGTTTCT 1317
1291 CAGCATGTACTTTTCT 1350
1318 CATCTGGCGGAGTGGCTTTCT 1377
1351 CATCT 1384
1378 TAAGCAATGGGCAACAGCCATTAAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTCAT 1437
1385 -----TCCAC 1389

QY 1438 TGAGAGAGAGAGCTGTATGACAGAGGCTGCAAGCTGAAGCATCTGACTAAGCATGATGC 1497
 Db 1390 ATACATGAAGAAATCTGACATTAAGTCCGCCAGAGCAAGAGAGCAAGCAAGCAAGATTC 1449
 QY 1498 CCAGCAGGATAGCTGGCTGAGCTCCACCTGAAGAGATTTGTTATTTGGCGTTCTAAT 1557
 Db 1450 CACTACCTATTAGATT-----GAAGTGGCAAGTGGGATGATCTGTCATGGCTTT 1503
 QY 1558 GCTGTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617
 Db 1504 CTTTCTCATCACTACACCTTTTATTCAACCTGGTGGTGAAGAGGCTGCTGCTGCTGCT 1563
 QY 1618 CTCATTTCTGCTGCTTCCACACAGTCAAGATGATCCGCAACATTTTACAGCATTT 1677
 Db 1564 GTCCATTTGATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 QY 1678 CAGAGAGGCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1737
 Db 1621 CCGAGAGCATATTATTGGCTTCTGCTAATACTTCCAGAGGATGCGAAGGCTCATGCTGCT 1680
 QY 1738 GTGGGATPACGGATACAGATGAGGGGAATGGCAACAGAGCAAGCTAGTGGATAATAA 1797
 Db 1681 GTGGGATPACGGATACAGATGAGGGGAATGGCAACAGAGCAAGCTAGTGGATAATAA 1740
 QY 1798 TACGTGGAAATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1857
 Db 1741 CACATGGAAATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 QY 1858 GTCCCTACCAATATTGATCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1917
 Db 1801 AGCTATGAGATCATGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 QY 1918 GATCGGCTATTCTGGGATGATATCAACAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1977
 Db 1861 CACTGGGTATTCTCTGATGATATCAACAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 QY 1978 ---AGAGCATCCCAAGGATATAGGAAAGCGATTACTTTACCGACCGGGTGAATTCAG 2034
 Db 1921 CACAGATACAGGCAACATATCAAGGAGATGACTATTATCTCCAACTGGGGAGTCCG 1980
 QY 2035 GTAGATCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2094
 Db 1981 TGTGACCGCTGAAGGTTCTCCAGTGTGCTCAACTGCTGCTCACTGATGATGTTACTTA 2040
 QY 2095 CAGATTCCGGGAAATGAAGTTGGACTACAGAGGCTCCATCTGGATATGATCGCACAGTAA 2154
 Db 2041 TCGCTTTGGACAGGTTTACACAGAGCCAGCGCTCTCCAGGCTTTTGAACGCTGTCGAA 2100
 QY 2155 GCGCTGATTTGGGAATAGGACTTCGATCTGACCTACCTGGAGGAGGCTTACACACAGA 2214
 Db 2101 TGCTGAGATTGGGAATAGGACTTTGAGCTTGATGCTGCTGGAGGAGGCTATACCACAGA 2160
 QY 2215 ACACCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2256
 Db 2161 ACATTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2202

RESULT 8

US-10-172-118-742
 ; Sequence 742, Application US/10172118
 ; Publication No. US20030224374A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dai, Hongyue
 ; APPLICANT: He, Yudong
 ; APPLICANT: Linsley, Peter
 ; APPLICANT: Mao, Mao
 ; APPLICANT: Roberts, Chris
 ; APPLICANT: Van 't Veer, Laura
 ; APPLICANT: Van de Vijver, Marc
 ; APPLICANT: Bernards, Rene
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
 ; FILE REFERENCE: 9301-175-999
 ; CURRENT APPLICATION NUMBER: US/10/172.118

; CURRENT FILING DATE: 2002-06-14
 ; PRIOR APPLICATION NUMBER: 60/380,770
 ; PRIOR FILING DATE: 2002-05-14
 ; NUMBER OF SEQ ID NOS: 2699
 ; SEQ ID NO 742
 ; LENGTH: 2472
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: NM_002219
 ; DATABASE ENTRY DATE: 2001-06-18
 ; US-10-172-118-742

Query Match 26.2%; Score 632.8; DB 15; Length 2472;
 Best Local Similarity 58.2%; Pred. No. 9.4e-184; Mismatches 802; Indels 93; Gaps 4;
 Matches 1247; Conservative 0;

QY 118 CAGCCTAATCACTTGGCCATCTGCTAATCGCTGGCTGGCCGGATTTCCCTCTCCCT 177
 Db 151 CACACCTTTGAAGCTTCTCAATCTGCTCAATGGCTGCTGATTATTCCTTCTCCACTGCT 210
 QY 178 CTTGCGCGTCATCCGTTTCGGAGTCGATTATCCATGAGTTTGCATCCGTGGTTCAACTACCG 237
 Db 211 GTTGTCTGCTGAGATTGAAGTTGAAGTTTATCAAGTTTGCATCCGTACTTTAATTATCG 270
 QY 238 GGCACCGCCTACATGGTGCAGAAATGGTTGGTGAACAATCTCTCACTGGTTGCAGAGCG 297
 Db 271 GACTACAGGTTCTGCTGGCTGAGGAGGGTTTTATAAAATTCATTAACCTGGTTTGTATGACCG 330
 QY 298 CGCATGCTATCGCTCGGCGAGGATTTGGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 357
 Db 331 AGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 390
 QY 358 GTCCGCGGAAATCCATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 417
 Db 391 CTCTGCT 450
 QY 418 CTGCT 477
 Db 451 CTGCT 510
 QY 478 CAAAGAGCT 537
 Db 511 CAAAGAGCTCAAGGATGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 570
 QY 538 TGCTACATCACTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
 Db 571 TGGATATATCTCCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 630
 QY 598 CTTGCGAGTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 657
 Db 631 CATGCTACTCACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 690
 QY 658 GGCCTGAGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
 Db 691 AGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 750
 QY 718 CATCAACCTGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777
 Db 751 GATCAACTTAATTCCTCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 810
 QY 778 TCTGCTGACAGCTACAGCACTTCTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 837
 Db 811 GATCATGTGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 870
 QY 838 CTTGCTGGGATTCGAACCGATACGCCAGTGAACACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
 Db 871 CTTTGTGGGTTTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 930
 QY 898 GCTCTCTTATGGCGCTGGCCACCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957
 Db 931 TCTCTGCCAGATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 990

QY 958 CCGAAGCTTTCATCGTCGGGATTTGCTGGTGGCGTGGCGTCTTTTGGCGCGT 1017
Db 991 TGAAGTTCCTTCCGGAGGCTCATCTCTGCTAGGCTTTGCTCTTCCACGCGGAGC 1050
QY 1018 GGTGCTCACCATGCTGGGCTTGTGGCCCGTGGAGTGGAGCTTCTACTGCTGGGA 1077
Db 1051 TCTCTCTATGCTGACAGGAAATATCTCCCTGGACGGGCTTTCTACTCAGTCTGGA 1110
QY 1078 TACTGCTACGCCAAGATCCATTCCTCATCTCCATCTCCGCTGGGAGCATCAGCCAC 1137
Db 1111 TCCCTCTTATGCTAAGAACCAATCCCATCTTCTCTCTGCTGAGCATCAGCCAC 1170
QY 1138 CACTGGTCTGCTTCTTCTTCTGATCTGCATCTGGTGTGGCGCTTCCAGTGGAGT 1197
Db 1171 AACCTGCTCTATCTATTTGACCTTGGAGCTCTCTCTCTCTCTCTCTCTCTCTCT 1230
QY 1198 GTGCTACTCATCAAGCATCAACGACGAGCGGCTTTTGGTGTGCTGTACGCCATCAG 1257
Db 1231 CTATTACTGCTTTAGCAACCTGCTGATGCGCGGATTTTATCATCATGATGCTGTGAC 1290
QY 1258 TCGGCTTACTTCTGCTGTGCTGATGCTGCTGTGATGCTGACCTCAGCGGCTGTG 1317
Db 1291 CAGCATGCTATTTTCTGCTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1350
QY 1318 CATGCTGCGGAGTGGCTTTTCTGCGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1377
Db 1351 CATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1384
QY 1378 TAAGCGAATGGCAGACCATTAAGCGCAGCCACCGAGTGTGAGTGTGAGTGTGAGT 1437
Db 1385 -----TCCAC 1389
QY 1438 TGAGAAGAACGCTGTACGACAGGCTGGCAAGCTGAAGCATCGTACTAAGCATGATC 1497
Db 1390 ATACATGAAGATCTGGACATTAAGTCCGACAGCAAGAGAGCAAGAGCAAGAGTTC 1449
QY 1498 CCAGCAGGATCTGGGCTCAGCTCCAACTGAGAGTATTTTATTTGGCGTCTTAA 1557
Db 1450 CACCTACCTATTAAGATTTT-----GAAGTGGCAAGTGGGATGATCTGCTGCTGCT 1503
QY 1558 GCTGTGTGATGCTTGTGCTTCCACTGCAAGTGGGTGCAAGCAATGCTACTCCAGTCC 1617
Db 1504 CTTTCTCATCACTACACCTTTTCACTCAAGTGGGTGACAGTGGGCTTCTCTCTCC 1563
QY 1618 CTCCTATGCTTGGCTTTCCACACAGTCAGATGCTCCGCAACATTTTAGACGATTT 1677
Db 1564 GTCCATTTGCTATCTGCTT-----GTGGTGGGATGGCAGTAGGATCATTTTGTAGCTT 1620
QY 1678 CAGAGGCTTACTACTGCTTTCCGAGAACACTGCGGATGATGCTGCGGTATGCTGTG 1737
Db 1621 CCGAGAAGCATATTTATGCTTCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
QY 1738 GTGGGATACGATACGAGATGGGATGGCAACAGACGAGCTAGTGGATATAA 1797
Db 1681 GTGGGATATGCTATCAGATTAAGCTATGGCAACGCAATTTTAGTGGCAATAA 1740
QY 1798 TAGGTGGAACATAGTCAATAGCTGCTGTTGGCAAGCAATGCTTCAACCGAGGAA 1857
Db 1741 CACATGGAATAATACCATTTTCTGAGTAGGCGAGCAATGGCTCCACAGAGGAAA 1800
QY 1858 GTCTACGAAATATGACATCTTGTAGTGGATGCTGCTGCTGCTGCTGCTGCTGCTG 1917
Db 1801 AGCTATGAGATCAGAGGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
QY 1918 GATCGGCTATTTGGCGATGATATCAACAGTCTCTGTGGATGCTGCGAATTTGTGAGG 1977
Db 1861 CACTGGGATTTCTCTGATGATATCAACAGTCTCTGTGGATGCTGCGGATTTGAGGAG 1920
QY 1978 ----AGAGCATCCAGGATTAAGGAGGATTTTACCAGCGCGCTGATTCAG 2034
Db 1921 CACAGATACAGGAAACATATCAAGAGGAGATGATTTATTTACTCCACTGGGAGTTCG 1980
QY 2035 GGTAGATCGGAGGTGCTCGGCGCTGCTCAACTGCTTATGATCAAAATTAAGCTACTA 2094

Db 1981 TGTGACCGTGAAGGTTTCTCCAGTGTCTCAACTGCTCATGTACAAGATGTGTACTA 2040
QY 2095 CAGATTCGGGAAATTGAAGTTGGACTACAGAGTCCATCTGGATATGATCGACAGTAA 2154
Db 2041 TCGCTTTGACAGGTTTACAGAGGCAAGCGTCTCCAGGCTTTGACCGTCTCGAAA 2100
QY 2155 CCGCTCATTTGGGAATAAGGATTCGATCTGATCTGACCTACTCGAGAGGCTTACACCA 2214
Db 2101 TCGTGAATTTGGGAATAAGGATTTGAGCTTTGATGCTCTCGAGGAGGCTTATACCA 2160
QY 2215 ACATGCGCTTGTTCGCTATCTATAGGTTGAAGAGCCGCTGAA 2256
Db 2161 ACATTTGGCTGCTCAGGATATACAGGTTAAGGACCTGGATAA 2202

RESULT 9

US-10-342-887-742
; Sequence 742, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Mao, Mao
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, René
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-742

Query Match 26.2%; Score 632.8; DB 16; Length 2472;
Best Local Similarity 58.2%; Pred. No. 9.4e-184;
Matches 1247; Conservative 0; Mismatches 802; Indels 93; Gaps 4;

QY 118 CAGCCTAATCACTTCGCGCATCTCTGTAATCGCTGCTGGCGGATTTTCTCTCGCCT 177
Db 151 CACACTTTTGAAGCTTCTCATCTGTCAATGGCTGCTGATATCTCTTCCACTGCTCT 210
QY 178 CTTCCGCTCATCCGTTTCGAGTCGATATCCATGAGTTTGTGCTGGTTCGAGTACCG 237
Db 211 GTTCTCTCTCTGAGATTTGAAGTGTATCCATGAGTTTGTGCTGGTTCGAGTACCG 270
QY 238 GCGCACCGCTTACATGTTGTCAGATGTTGTGTAACAATCTCTCAACTGTTCCAGGCG 297
Db 271 GACTACCAAGTTCCTGGCTGAGGAGGCTTTTATATTCATTAATCCATGCTGATGACCG 330
QY 298 CGATGATTCGCTCGGAGGATTTGGCGGTGACGCTCTATCCCGCTCATGATGATAC 357
Db 331 AGCTGTGTACCTTTGGGACGAATCATTTGGAGGAACAATTTATCCAGGTTTATGATCAC 390
QY 358 GTCCCGGCGAATTCATTTGGCTGTGACGCTACTCAACATACCGGTCCATATTCGTGACAT 417
Db 391 CTCGTCTCATCTCACTACCATGCTACTCTCTTTTCCATCACCATCGACTTCGATGT 450
QY 418 CTTGCTGTTCCTGGCGGATTTTCAGTGGCTGACCTCCATTCACCTACCTGCTGAC 477
Db 451 CTTGTGTGTTCTGGCGCT 510

156 CACACTTCAAAGCTTCTCATCTCTCGATGGCTGCTGTGTTATCTTTTCTACTGCTCT 215
178 CTTGCGCGTATCCCGTTTGGATCGATATCATAGTTTGAATCCGTTGTTTCACTACCG 237
216 TTTTGTGCTGTGAGATTGAAGGTGTCATCCATGAGTTTGCATCCGTACTTAAATATCG 275
238 GGCACCGGCTACATGGTGCAGATGGTTGTACAACTTCTCAACTGGTTTGCACGAGCG 297
276 GACTACCGGTTTCTGGCTGAGGAGGGTTTATAATTCATTAACCTGGTTTGTAGACCG 335
298 CGCATGGTATCCGCTCCGACGATTTGGCGGTACCGTCTATCCCGGCTGTATGATTAC 357
336 GCGTTGGTACCTTTTGGGCGGAATCATTTGAGAGAACAAATTTACCCAGGTTTAATGATCAC 395
358 GTCCGGCGGAATCCATTTGCTGCTGACGTACTCAAGATACCGTCCATATTCGTGACAT 417
396 TTCTGCTGCAATCTACCATGTACTCCATTTCTCATATCAATTTGACATTTGGATGT 455
418 CTGCGTTTCTGGGCGGATCTTCAAGTGGGCTCACTCCATCTCCACCTACCTGTGTGAC 477
456 CTGTGTTTCTGGGCGGCTTCTCTCTTTTCCACCACCATCGTTACGTACCACCTTAC 515
478 CAAGAGCTGTGGTCCGGGCGGCGCTCTTCCGCGCAGGTTTATCCGCTATCCGCTGCG 537
516 CAAGAGCTCAAGGATGAGAGCTGGGCTTCTTCTGCTGCCATGATGCTGTAGTTC 575
538 TGGTACATCATAGTGGTGGGCTGGATCTGATCAAGAGGGCATTCGCCATATTCGC 597
576 TGGGTATATTTCTGATCTGTAGTGGCTCTATGATATAATGAAGAAATTCATCTTTTG 635
598 CTGAGCTTCACTACTTCTCTGGTGGCTGCTGATGAGAGCTGATCCGCTGTCTGCTG 657
636 CATGCTGCTTACTTACTATCATGTGATCAAGGCTGAGAGCTGTTCCATCTATTGGCG 695
658 GGCACGACCGCTTGTCTTACTTCTACATGGTGTCCGCTGGGCTGGCTGCTGCTGCTCAT 717
696 TGCCAAAGTGTGCGCTCGCTTATTTCTACATGGTCTCTTTCATGGGAGGCTATGTGTCCT 755
718 CATCACTGTATACCCCTGACGCTCTGCTGATGCTGATGAGAGCTGATCCGCTGCTG 777
756 GATCACTGTATCTCTTACATGCTGCTGCTGATGCTGATGAGAGCTGCTTCTTCTCACCG 815
778 TCTGTGACGAGCTACAGCACTTCTACATCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCT 837
816 GATCTACGTAGCTACTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 875
838 CTTCTGGGATTCACACGATACGACAGCTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
876 CTTTGTGTTTCCAGCCGCTCTTCTATGAGACATGAGGCTTCTGAGTGTGCTGCTGCTGCT 935
898 GCTCCTTATGGCGGTGGCCACTTGGCCATTTGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957
936 TCTCTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 995
958 CCGGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
996 CGAAGTCTTTTCCGAGTGTATCTCCCTGGTGGCTTGTCTCTCTCTCTCTCTCTCTCTCTCT 1055
1018 GGTGCTCAACATGCTGGGCTGTTGGCCCGTGGAGTGGAGCTTCTTCTCTCTCTCTCTCTCTCT 1077
1056 TCTCTCTATGCTAACAGGAAATTTCTCCCTGACAGGCGGCTTCTTCTCTCTCTCTCTCTCT 1115
1078 TACTGGCTACCGAGATCCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1137
1116 TCCCTCTTATGTAAGATAATATCCCATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1175
1138 CACTTGGTCTGCTTCTTCTTGTATGCTGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
1176 AACCTGGTCTTCTTACTATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
1198 GTGGTACTGCTACAGCAGATCAACGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257

1236 CTATTACTGCTTTAGCAACCTGTGTGATGCTCGGATTTTTTATCATCATGATGATGCTGTGAC 1295
1258 TCGGTTTACTTCCGCTGGTGTGATGGTGGCTTTGATGTTGACCTCAOCGGTGTGTG 1317
1296 CAGCATGTACTTTTTCAGCTGTAATGGTGGCTCTAAATGCTGATGTCACCTGTTATG 1355
1318 CATGTCGCGGAGTGGCCTTTTCGGGACTGTTGGATGTTGCTTCTGCAAGAGGATTCGTC 1377
1356 CATCTTTCTGGCATTTGGTGTTCAGGCTGCTGTCCACATATATGAAAAA----- 1406
1378 TAAGCAATGGGCACGCCATTAAGCGCAGCCACCGAAGTGGATGGAAGTGAAGGATTCAT 1437
1407 ----- 1406
1438 TGAGAAGAGACGCTGTACGACAAGGCTGGCAAGCTGGAACATCGTACTAAGCATGATGC 1497
1407 TCTGGACATAAGTCCCGACAGACAAGAGACAGACAGATTTACTTACCTAT 1466
1498 CCAGCAGATACTGGCTCAGCTCCAACTGAAAGATTTGTTATTTTGGCGCTTCTAAT 1557
1457 TAAGAAATGAGG-----TGGCGAGTGGGATGATACTGCTCATGGCTTT 1508
1558 GCTGTTGATGATGTTGCTGTCCACTGCACTGCGTGGGTGACCAAGTCCCTACTCCAGTGC 1617
1509 TTTTCTCATCACCTACAGGTTTCAATCGATTTGGGTGACCGATGGAAGCTTATCTTCTCC 1568
1618 CTCCATTTGCTGCTTCCACAACTCAAGATGGAATCCCGCAACATTTTAGACGATTT 1677
1569 CTCCATTTGCTGCTGCTC-----GTGGTGGGATGCGAGTAGGATCAATTTTGTGACAT 1625
1678 CAGAGAGCTTACTTACTGGCTTTTCGCAAGACACTGCGGATGATGCTGCGCTTATGCTTG 1737
1626 CCGAGAGCGTATTTATTTGGCTCCGTCAATCTCCAGAGATGCAAAAGTCAATGCTG 1685
1738 GTGGATTAOCGATACAGATAGCGGAATGGCAACAGAACCGCTAGTGGATATAA 1797
1686 GTGGGATTTAGCTTACCAATTTACTGCAATGGCAAAATCGGACAAATTTTAGTGACAATA 1745
1798 TACGTGGAACAAATGATCACATAGCGTGGTGGCAAGCAATCTTCAACCGAGAGAA 1857
1746 CACATGGAATATACCATATTTCTCGAGTAGGCGGCAATGGCATCCACAGAGAAA 1805
1858 GTCTACGAAATATGATCTCTTGACGTGGAATAGTTTGGTGGTATCTTTGGCGGTGT 1917
1806 AGCCTATGAAATCATGAGGAGCTTTGATGTCAGTATGCTGCTGCTGCTGCTGCTGCTGCT 1865
1918 GATCGGCTTATTCGGCGATGATATCAAGATTTCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 1977
1866 TACTGGGTATTTCTCGGATGATATCAAGATTTCTTGGATGCTGCTGCTGCTGCTGCTGCTGCT 1925
1978 ---AGAGCATCCCAAGCAATTAAGGAAGCAATTTACTTACCGACCGGCTGAAATTCAG 2034
1926 CACAGACAGAGAGACACATTAAGGAGATGACTACTATCTTCTTCTTCTTCTTCTTCTTCTTCT 1985
2035 GGTAGATCCGAGAGCTCCGCGCTGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2094
1986 TGTGATGCTGAGGTTTCTCGGCTGCTGCTCAACTGCTTATGATCAAAATGTTTACTTA 2045
2095 CAGATTCCGGGAATTAAGTTGGAATACAGAGTCCATCTGATATGATGCTGCAACGTAA 2154
2046 CCGCTTTGGGCGGCTTACACAGAGCAAGCGCTCCACAGGCTTTGACCGCTTTCGAA 2105
2155 CGCGCTCATTTGGGAATGAGCTTCCGATCTGACCTACCTGAGGAGGCTTACACACAGA 2214
2106 TGCTGAGATTTGTAATTAAGCTTTGAGCTTGTATGCTGAGGAGGCTTATACACAGA 2165
2215 AACTGCTGTTGCTGCTGCTTATAGGTTGAAGAGCGGCTATGA 2256
2166 AACTGCTAGTACAGGATATACAGGTAAAGGACCTGGATAA 2207

RESULT 11
US-10-028-384-9

; Sequence 9, Application US/10028384
; Publication No. US2003014285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 9
; LENGTH: 3094
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_008408
; DATABASE ENTRY DATE: 2000-11-01
; RELEVANT RESIDUES: (1)..(3094)
US-10-028-384-9

Query Match 26.1%; Score 631.2; DB 15; Length 3094;
Best Local Similarity 58.2%; Pred. No. 3 3e-183;
Matches 1246; Conservative 0; Mismatches 803; Indels 93; Gaps 4;
QY 118 CAGCCTAATCACCTTCGCCATCTCTGTAATCGCTGGCTGGCGGGAATTTCTCGCCT 177
DB 156 CACACTTCTAAAGCTTCTCATCTCTGTCGATGGCTGTGTATCTTTTCTACTCGTCT 215
QY 178 CTTCCCGCTCATCCGTTTCCAGTCGATTATCCATGAGTTTGTATCGTGGTTCACATCCG 237
DB 216 TTTTGTCTGTGAGATTGAAAGTGTCAATCCATGAGTTTGTATCGTACTTTAATATCG 275
QY 238 GGCACCCGCTTACATGTTGAGAAATGTTGGTACAACTCTTCACTCACTGTTCCGAGCG 297
DB 276 GACTACCCGCTTCTGCTGAGGAGGGGTTTATAAATCCATACTGTTTGTATGACCG 335
QY 298 CGCATGTTATCCGCTCGGCGAGGATTGGGCGGTACCGTCTATCCGGGCTCATGATTAC 357
DB 336 GCCTTGTACCTTTGGGCGGATCAITGGAGAAACAATTTACCAGGTTTAAATGATCAC 395
QY 358 GTCGGCGGAATCAATGTTGCTGTGACGATCTCAACATACCGGTTCCATTTGCGACAT 417
DB 396 TTCTGCTGCAATCAACATGATCTCCATTTCTTCCATATCACTAATGACATTCGGAATGT 455
QY 418 CTGCGTGTCTGCGCGCGATCTTCACTGTCCTGACCTCCATCTCCACCTACCTCTGAC 477
DB 456 CTGTTTCTGCGCGCGATCTTCTCTCTCTTTCACCACTCTTACGTACCACTTAC 515
QY 478 CAAGAGCTGTGTCGCGCGCGCGCTCTTTCGCGCGCGAGTTGATCGCCATCTGTC 537
DB 516 CAAGAGCTCAAGGATCAGGAGCTGGGCTTCTTGTGTCGATGATGCTGTGATTTCC 575
QY 538 TGGCTACATCAGTAGTGGTGGTGTGATCGTACGATTAACGAGGCAATGCCATATCGC 597
DB 576 TGGGTATTTCTGATCTGTAGTGGTCTCTATGATATGAGGAAATGCTATCTTTTG 635
QY 598 CTTGAGTTACCTTCTCTGTTGGTGGCTCAGTGAAGACTGGAATCGGTTCTGCTG 657
DB 636 CATGCTGCTTACTTACTACATGTTGATCAAGGAGTGAAGCTGTTTCCATCTATTGGGC 695
QY 658 GCGCGAGCGCTTGTCTTACTTCTACATGTTGTGCGCTGGGCTGAGTCTGTTCTAT 717
DB 696 TGCCAAGTGTGCGCTTATTTCTACATGTTCTTCTATGGGAGGCTATGTTCT 755
QY 718 CATCAACCTGATACCCCTGACGTTCTGCTGCTCATATGAGGAGTACTCCCGCG 777
DB 756 GATCAACTTGTATCTCTACATGTTGCTGCTGAATGCTGACAGGCGGTTTCTCACCG 815
QY 778 TCTGTGACGAGCTACAGCACTTCTACATCTCTGGGAGCTGCTGTTCTCCATCGATCCC 837
DB 816 GATCTACGCTACTGTACTGTTTACTGCTGCGGCGACCACTTCTTCTATCAGATTTC 875

QY 838 CTTCTGTTGGAATTCACCGATACGACACAGTGAACACATGCTGCTGGAGTGTGT 897
DB 876 CTTTGTGTTTCCAGCCGCTCTTTTCATCAGAACACATGCGACCTTTTGGAGTGTGTG 935
QY 898 GCTCTTATGCGCGTGGCCACCTTTGCGCATTTTCAGTCCGTGCTGCGGCAACGAGTT 957
DB 936 TCTCTGTAGATCCATGCTTTGTTAGATTACCTGCGAGCAAGTTGAATCCACAGCAAT 995
QY 958 CCGAAGCTGTTCATCGTGGCGGATTCCTGTTGGGCTTGGGCTCTTTGTCGCCGTCT 1017
DB 996 CGAAGTCTTTTCCGAGGTATCTCCCTGGTGGTCTTGTCTCTCTCACTGTGGAGC 1055
QY 1018 GGTGCTCACCATGCTGGGCGTGTGGCCCGTGGAGTGGAGCGCTTCTACTCGCTGGGA 1077
DB 1056 TCTCCTCATGCTAACAGGAAATTTCTCCCTGGACAGGGGCTTTCTACTCTCTGCTGA 1115
QY 1078 TACTGCTAGCCAGATCCACATTCCTCATTCATTCGATCCGTTGCGAGCATCAGCCAC 1137
DB 1116 TCCCTCTTATGTAAGAATAACATTCCTCATTTTGCATCTGTTTCTGAGCACAGCCAC 1175
QY 1138 CACTTGTCTCTGTTCTTCTTGTATCTGCACATCTGTTGTCGCTTCCAGTGGAGT 1197
DB 1176 AACTTGTCTTCTCTACTATTTGATCTACAGTCTCTGTTCTCATGTTTCCAGTGGCCT 1235
QY 1198 GTGCTACTCATCAGCAGATCAGCAGAGCGGCTTTGCTGCTGTGCTGCTGACCCATCAG 1257
DB 1236 CTATTACTGCTTTAGCACTGTTCTGATGCTGCGGATTTTATCATCATGATGTTGCTGAC 1295
QY 1258 TCGGCTTACTTCTGCTGTTGATGCTGCTGTTGATGTTGACCTCCAGCGGCTGCTG 1317
DB 1296 CAGCATGTTACTTTTCTGCTGTTAAATGCTGCTGTTAAATGCTGCTGCTGTTG 1355
QY 1318 CATGCTGCGGAGTGGGCTTTTCCGAGCTGTTGAGTGTCTGCTGCAAGAGATTCGTC 1377
DB 1356 CATTCTTCTGCTGATGTTGTTTCCAGGCTGTTCCACATATATGAAAAA----- 1406
QY 1378 TAAGCAATGGGACACAGCATAAAGCGCAGCCACCAAGTGGATGAGCTGAGGATTCAT 1437
DB 1407 ----- 1406
QY 1438 TGAGAAGAGAGCTGTAGCAAGCTGGCAAGCTGAGCATCGTACTAAGCATGATGC 1497
DB 1407 TCTGGACATAGTCTCCCGACAGACAGAGCAAGCAAGCAAGATTTCTACTTACCCAT 1466
QY 1498 CCAGCAGGATGCTGGGCTGAGCTCCACCTCAAGAGATTTGTTATTTGGCGCTTTCTAAT 1557
DB 1467 TAAGATGAGG-----TGGGAGTGGGATGATGATGCTGCTGCTGCTT 1508
QY 1558 GCTGTTGATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617
DB 1509 TTTTCTCATCAGCTACAGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1568
QY 1618 CTTCAATGCTTGGGCTTTCCACCAAGTCAAGATGAGTCCCGCAACATTTTAGACGATTT 1677
DB 1569 CTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1625
QY 1678 CAGAGGCTTACTGCTGCTTTCGAGAACACTGCGGATGATGCTGCTGCTGCTGCTGCTGCT 1737
DB 1626 CCGAAGGCTGATTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1685
QY 1738 GTGGATACCGATACCGAGTAGCGGGAATGGCAACAGAACAGAGCTAGTGGATATAA 1797
DB 1686 GTGGATATGCTTACCAATTTTCTGCAATGGCAATCGGACAAATTTTAGTGGACATAA 1745
QY 1798 TAGTGGACATAGTACATAGCTGTTGGCAAGCATGCTTCAACCGAGGAA 1857
DB 1746 CACATGGAATATACCATATTTCTGAGTAGGCGAGCAATGCAATCCACAGAGAAA 1805
QY 1858 GTCCTACGAAATATGACATCTTTGAGCTGAGCTAGCTTTTGGTGTCTTTGGCGGT 1917
DB 1906 AGCTATGAATCATGAGGAGCTTGTGTCAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1865

QY 1918 GATCGCTATTCTGGGATGATATCAACAGTTCTCTGTGGATGGTCCGAAATTCGTAGGG 1977
Db 1866 TACTGGGTATCTTCGGATGATATCAACAGTTCTCTGTGGATGGTCCGAAATTCGTAGGG 1925
QY 1978 ---AGAGCATCCAAAGGACATTAAGGAAAGGATTAATTTACCGACCGCGGTGAATTCAG 2034
Db 1926 CACAGAGACAGGAAGACACATTAAGGAAAGGATTAATTTACCGACCGCGGTGAATTCAG 1985
QY 2035 GGTAGATGCCGAAGGTCTCCGCGCTCTGCTCAACTGCTTATGTACAAATTAAGCTACTA 2094
Db 1986 TGTGTATCTGTAGGGTCTTCGGGTCTGCTCAACTGCTTATGTACAAATTAAGCTACTA 2045
QY 2095 CAGATTCCGGGAAATTAAGTTGGACTACAGAGGTCATCTGGATATGATCCACACGTAA 2154
Db 2046 CGCTTTGGGAGGCTACACAGAAAGCAAGCGTCCACCGCTTTGACCGTGTTCGAA 2105
QY 2155 CGCGCTCATGGAAATAGGATCTCGCTCAACTGCTTATGTACAAATTAAGCTACTA 2214
Db 2106 TGTGTAGATTGGTAATAAGACTTTGAGCTTGATGCTCGGAGGAGGTATACACAGA 2165
QY 2215 AACTGGCTTGTTCGATCTATAGGTGAAGAGCGGCATGA 2256
Db 2166 AACTGGCTTGTTCGATCTATAGGTGAAGAGCGGCATGA 2207

RESULT 12

US-10-417-375-99
; Sequence 99, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 5404
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-99

Query Match 26.0%; Score 628.6; DB 18; Length 5404;
Best Local Similarity 58.3%; Pred. No. 2,7e-182;
Matches 1238; Conservative 0; Mismatches 794; Indels 93; Gaps 4;

QY 118 CAGCCTAATCAGCTTCGCCATCTGCTAATCGCCTGGCTGGCGGATTTCTCTCCCT 177
Db 180 CACACTTTTGAAGCTTCTCAATCTGTCAATGGCTGCTGTATATATCTCTCCACTGCT 239
QY 178 CTTCGCGCTCATCCGTTTCGAGTCGATTATCATGAGTTTGTATCGGTGTTCAACTACCG 237
Db 240 GTTGTGTCCTGAGATTGAAGTGTATCATGAGTTTGTATCGGTGTTCAACTACCG 299
QY 238 GGCACCGCTTACATGCTGTCAGAAATGGTGTGTAACCTTCTCACTGGTTCGACGAGCG 297
Db 300 GACTACAGGTTCCCTGGCTGAGAGGGGTTTATTAATTTCCATTAATCTGTTGATACCG 359
QY 298 CGCATGGTATCCGCTCGGAGGATTTGGGGGTATCCGCTTATCCCGGCTTGAATATAC 357
Db 360 AGCTGGTACCTTTGGGAGCAATCATTTGGAGCAATTTACCGAGTTTAAATGATCAC 419
QY 358 GTCCGCGGAATTCATGCTGCTGACGATCAATCAATACCGGTTCATATTCGTGACAT 417
Db 420 CTCTGCTGCAATCTACCATGTACTCTCAATTTTTCACATCATCCATCGACATTCGGAATGT 479
QY 418 CTGCGGTTCCTGGGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGAC 477
Db 480 CTGTGTGTTCTGGGCGGCT 539
QY 478 CAAGAGAGCTGTGGTCCGCGGGCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 537

Db 540 CAAGAGAGCTCAAGAGATCAGGGGCTGGGCTTTCTGTGCTGCCAATGATTTCTGTAGTTC 599
QY 538 TGGCTACATCAGTATAGTTCGGTGGATCGTACGATAACGAGGCAATTCCTCATTTCCG 597
Db 600 TGAATATATCTCCGATCTGTGGCTGCTCTATGATAATGAAGGATTCCTCATTTTG 659
QY 598 CTTGAGTTCACCTACTACTCTCTGCTGGTGGCTCAGTGAAGACTGGATCCGTTCTGGTC 657
Db 660 CATGCTACTACCTACTACATGTGATCAAGGAGTAAAGACTGTTCCATCTGTGGG 719
QY 658 GGCAGCAGCCGCTTGTCTCTTCTATCTCTGCTCCGCTGGGCTGGCTGCTGCTGCT 717
Db 720 AGCTAAGTGTGGCTTGTCTTATTTCTATCTGCTCTGCTCAATGGAGGTTATGCTGCT 779
QY 718 CATCAACCTGATACCCCTGACGCTTCTGCTACTGCTCATTTATGGGAGGTTACTCGCGG 777
Db 780 GATCAACTTAATCTCTCCACGCTCTGCTGCTGCTCATGCTCACAGGCGCTTTCTCACCG 839
QY 778 TCTGCTGACAGCTACAGCACCCTTCTATCTCTGGGACTGCTGTTCTCTCAATGAGATCCC 837
Db 840 GATCTATGTGGCTACTGCTACTGTTTACTGCTGGCACTATCTTCTATGAGATCTC 899
QY 838 CTTGCTGGGATTCGAACCGATACGACAGTGAACACATGGCTGGCTGGGAGTGTGT 897
Db 900 CTTTGTGGGTTTCAGCCTGCTCTTTCATGAGACATGAGCAGCTTTGGGCTTTGG 959
QY 898 GCTCTTATGCGCGTGGCCACCTTTGCGCCATTTTGCAGTCTGCTGCTGCGCAACGAGTT 957
Db 960 TCTCTGCCAGATCATGCTCTTTGTGGATTACCTGCCAGCAAGTTGAATCACACAAT 1019
QY 958 CCGGAGAGCTGTTTCATGCTGGGAGTTCCTGCTGGGCGTGGGCTCTTTTGGCGCTGCT 1017
Db 1020 TGAAGTCTTTTCGAGCGTCTCTCTGTTAGGCTTTGCTCTCTCTCTCTCTCTCTCTCT 1079
QY 1018 GGTGCTCACCATGCTGGCGCTTGTGGCGCGTGGAGTGAACGCTTCTCTCTCTCTCTCT 1077
Db 1080 TCTCTCTCATGCTGACAGGAAAAATATCTCTCTGACGCGGCTTTCTACTGCTGCTGA 1139
QY 1078 TACTGCTACGCCAGATCCCATTTCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1137
Db 1140 TCCCTCTTATGCTAAGAACAAATCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1199
QY 1138 CACTTGGTTCGCTTCTCTCTTGTATCTGACATCTCTGCTGCTGCTGCTGCTGCTGCT 1197
Db 1200 AACCTGGTCTCATATCTATTTTGACCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1259
QY 1198 GTGGTACTGCTCAAGCAGATCAAGCAGCGGCTTTCTGCTGCTGCTGCTGCTGCTGCT 1257
Db 1260 CTATTAAGTCTTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1319
QY 1258 TGGGCTTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
Db 1320 CAGCATGCTATTTTCAGCTGTAATGGTGGCTTAATGCTAGTGTGGCAGCTGTTATGTG 1379
QY 1318 CATGCTGGCGGAGTGGCTTTTTCGGGACTGTTGGATGCTGCTGCTGCTGCTGCTGCTGCT 1377
Db 1380 CATCTCTCTGCGCATTTGAGTCTCCAGGTCCTG----- 1413
QY 1378 TAAGGAAATGGGCAACAGCATAAGCGCAGCCACGAAAGTGAATGAAGCTGAGGATTCAT 1437
Db 1414 -----TCCAC 1418
QY 1438 TGAGAAGAGAGCTGTACGACAGGCTGGCAAGCTGAAGCATGCTACTAAGCATGATGC 1497
Db 1419 ATACATGAAGAAATCTGGACATAAGTCTGTCAGACAAGAGAGCAAGAGCAAGGATTC 1478
QY 1498 CCAGCAGGATATCTGGCGTCACTCCAACTGAAGAGTATTTGTTATTTTGGCCGCTTCTAAT 1557
Db 1479 CACCTACCTATT-----AAGAAATGAAGTGGCAAGTGGGATGATCTGCTCATGGCTTT 1532
QY 1558 GCTGTGTGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1517

Db 1533 CTTTCTCATACCTACACCTTTTCATTCAACCTGGTGACAGTGAGGCTACTCTTCTCC 1592
 Qy 1618 CTCGATTTGCTTGGCTTTCACACAGTCAAGATCGATCCCGCAACATTTTACAGATTT 1677
 Db 1593 GTCCATTTACTATCTGCCCC---GTGGTGGGGATGGCAGTAGGATCATATTTGATGACTT 1649
 Qy 1678 CAGAGAGCTTACTACTGCTTTCGACAGAACACTGCGATGATGCTCGGCTTATGCTTG 1737
 Db 1650 CCGAGAAGCATATTTATGGCTTCCTCATATACTCCAGAGGATGCGAAGTCAATGCTCTG 1709
 Qy 1738 GTGGATTTACGATACCGATAGCGGAATGGCAACAGAACGACCTAGTGATTAATA 1797
 Db 1710 GTGGATTTATGGCTATACAGATTAAGCTATGGCAACCGAACAATTTTATGAGCAATAA 1769
 Qy 1798 TACGTGGAAACAATAGTACATAGGCTTGGTGGCAAGGCAATGCTTCAACCGAGAGAA 1857
 Db 1770 CACATGGAATATACCCATATTTCTCGATAGGCGAGGCAATGGCGTCCACAGAGGAAA 1829
 Qy 1858 GTCTACGAAATATGATCTCTTGAGCTGAGTACGTTTGGTGATCTTTGGCGGTGT 1917
 Db 1830 AGCCTATGAGATCATGAGGAGCTCGATGTGAGTATGCTGCTGCTCAATTTTGGAGGCT 1889
 Qy 1918 GATCGGCTATTCTGCGGATGATCAACAAGTTCTTGGATGGTCCGAAATGCTGAGGG 1977
 Db 1890 CACTGGGTATTTCTCTGATGATCAACAAGTTTCTTGGATGGTCCGGAATGGAGGAG 1949
 Qy 1978 ---AGAGCATCCGAGGCAATTAAGGAAGGATTTTACCGACGGGGTGAATTGAG 2034
 Db 1950 CACAGATACGCGCAACATATCAAGGAGATGACTATTATPACTCCAACCTGGGAGTTCCG 2009
 Qy 2035 GGTAGATCCGAAGGTGCTCCGGGCTGCTCAACTGCTTATGTACAAATTAAGCTACTA 2094
 Db 2010 TGTGACCGGTGAAGTTCTCAGTGTGCTCACTGCTCATGTACAAGATGTTACTA 2069
 Qy 2095 CAGATTCGGGAATGAAGTGGACTACAGAGGTCCATCTGGATATGATCGCACAGTAA 2154
 Db 2070 TCGCTTTGGACAGGTTTACAGAGCAAGCGCTCCAGGCTTTGACCGGTGTCGAAA 2129
 Qy 2155 CGCGCTCATTTGGGAATAGGACTTCGATCTGACCTACCTGGAGAGGCTTACACACAGAA 2214
 Db 2130 TGCTGAGATTTGGGAATAGACTTTGAGCTTGATGCTTGGAGGAGATATACACAGAA 2189
 Qy 2215 ACAGTGGCTTTTGGCATCTATAG 2239
 Db 2190 ACATTTGGCTGTGAGGATATACAG 2214

RESULT 13

US-10-417-375-97
 ; Sequence 97, Application US/10417375
 ; Publication No. US20040219528A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; APPLICANT: Marc Malandro
 ; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
 ; FILE REFERENCE: 529452001600
 ; CURRENT APPLICATION NUMBER: US/10/417,375
 ; NUMBER OF SEQ ID NOS: 176
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 97
 ; LENGTH: 5827
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-417-375-97

Query Match 26.0%; Score 628.6; DB 18; Length 5827;
 Best Local Similarity 58.3%; Pred. No. 2.8e-182;
 Matches 1238; Conserved 794; Indels 93; Gaps 4;
 Qy 118 CAGCCTAATACCTTCGCGATCTGCTAATCGGCTGGCTGGCGGATTTTCTCTCGCT 177
 Db 180 CACACTTTGAAGCTTCTCATCTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 239

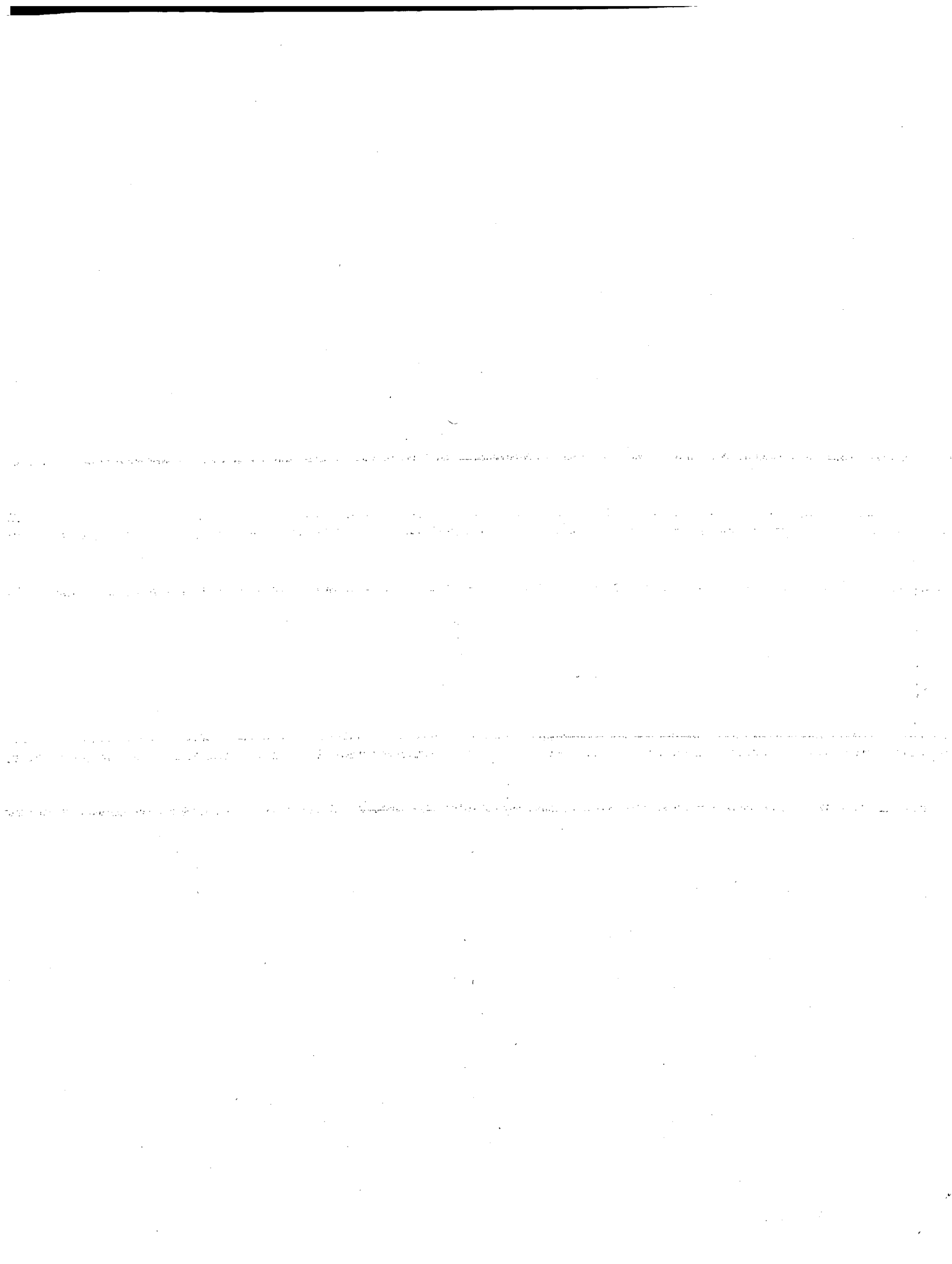
Qy 178 CTTGCGGCTCATCCGTTTCGAGTCGATTTATCATGAGTTTGTATCGGTGGTTCAACTACCG 237
 Db 240 GTTTGCTGCTCGAGATTTTGAAGTTTATCCATGAGTTTGTATCGGTACCTTTAATATTCG 299
 Qy 238 GGCACCGGCTCATCATGCTGAGATTTGGTGTACAACTTCTCTCACTGGTTTCGACGAGCG 297
 Db 300 GACTACCAAGTTCTCTGCTGAGAGGGTTTATATAATTTCCATAACTGGTTTGTATGACCG 359
 Qy 298 CGCATGGTATCCGCTCGGAGGATTTGTGGCGGTACCTCTATCCCGGCTGATGATTAC 357
 Db 360 AGCTGGTACCTTTTGGGAGCAATCATTTGAGGAGCAATTTTACCAGGTTTAAATGATCAC 419
 Qy 358 GTCCGCGGAATCCATTTGGCTGCTGACGATCTCAACATACCGGTCCCATATTTCTGTGACAT 417
 Db 420 CTCTGCTGCAATCTACCATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGT 479
 Qy 418 CTGCTGTTCTTGGGCGCGATCTTCAGTGGCTGACCTTCCATCTCCACCTACCTGCTGAC 477
 Db 480 CTGTGTTTCTTGGGCGCGCTCTCTTCTCTCTCTTCCACCATCGTCACTACCTTAC 539
 Qy 478 CAAGAGCTGTGTGCTCGGCGCGCGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537
 Db 540 CAAGAGCTCAAGGATGACAGGGCTTGGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 599
 Qy 538 TGCTACATCAGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
 Db 600 TGGATATATCTCCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659
 Qy 598 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 657
 Db 660 CATGCTACTCACTACTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719
 Qy 658 GSCGCGAGCGGCTTGTCTACTTCTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
 Db 720 AGTAAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779
 Qy 718 CATCAACCTGATACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777
 Db 780 GATCAACTTAATTTCTCTCCAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 839
 Qy 778 TGTGCTGACGAGTACAGACCTTCTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 837
 Db 840 GATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 899
 Qy 838 CTTGCTGGGATTTCCAAACCGATACGACCAAGTGAACACATGCTGCTGCTGCTGCTGCTGCT 897
 Db 900 CTTTGTGGGTTTCCAGCTGCTCTTTCATGAGACACATGCTGCTGCTGCTGCTGCTGCTGCT 959
 Qy 898 GCTCCTTATGCGGCTGCGGACCTTGGCGGATTTGAGTCCGCTGCTGCTGCTGCTGCTGCTGCT 957
 Db 960 TCTCTCCAGATCTCATGCTTGTGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1019
 Qy 958 CCGGAAGCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
 Db 1020 TGAAGTTCTTTCGAGAGGCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1079
 Qy 1018 GTGCTCAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077
 Db 1080 TCTCCTCATGCTGACAGGAAAATATCTCCCTGGAGGGGGCTTCTACTGCTGCTGCTGCT 1139
 Qy 1078 TACTGGCTACGCAAGATCCATTTCCCATCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1137
 Db 1140 TCCCTCTTATGCTAAGAACACATCCCATCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1199
 Qy 1138 CACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
 Db 1200 AACTGGCTCTCATATTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259
 Qy 1198 GTGGTACTCATCAAGAGATCAACGAGCGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257
 Db 1260 CTATTACTGTTAGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1319

QY 899 CTCCTTATGGCGGTGGCCACCTTGGGCCATTTGCAAGTCGGTGTCTGTGGCGCAACGAGTTTC 958
DB 1006 CTTGTACAGCTGATCGGATTTCGCGAAGTCGTCCGACGACTCGTGTGGCTGGCGAAGCAATTC 1065
QY 959 CGGAAGCTGTTTCATCGTGGCGGATTTGCTGTGGCGGTGGCGCTCTTTGGGCGGTCG 1018
DB 1066 CAGCTCCTTCTCAAGCTTTTGTGCTGGCGGATTTCTGCTCGATTTTCTGCTGCTGCTG 1125
QY 1019 GTGCTCACACGCTGCTGGCGGTTTGGGCGCGTGGAGTGGACGCTTCTACTGCTGTGGAT 1078
DB 1126 ACTTTGACTTTCTCTGGATGATCGCCCTTCTGCTGGAAGATTTTATTTCTTTGGGAT 1185
QY 1079 ACTGCTAGCGCAAGATCCATCCCATCATTCATCCGTTGCGAGCATCAGCCACC 1138
DB 1186 ACTGCTATGCGAAGTCCCATGCCCATTTATGCTCCGCTCTCCGACACAGCCACC 1245
QY 1139 ACTTGGTCTGTTCTTTCTTTGATCTGCACATCTGTGTGGCGCTTCCAGTGGAGTG 1198
DB 1246 GCTTGGCGCTCATTTCTACTTTGACCTCAGATGCTTATCTTTTCTTTTCTGCTGCGGTGTC 1305
QY 1199 TGGTACTGCATCAAGCAGATCAACGAGCGCGCTTTTCTGCTGTGTCAGCCATCAGT 1258
DB 1306 TCTGCTGTTTCAGGAGCTTCGCGATGAGCAGATCTTCAATCATATTTATGCGGTCTC 1365
QY 1259 GGGTTTACTTCTGCTGTGATGCTGCTGTTGATGTTGACCTCTACGCGGTGCTGCTG 1318
DB 1366 AGTGCCTATTTTGGCGGTGCTGATGTTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 1425
QY 1319 ATGCTGGCGGAGTGGCGCTTTTGGGACTGTTGATGTTTCTGCAAGAGATTCGCT 1378
DB 1426 GTTCTCTCGCAATGCTTCTCCAACTTCTCGAGCGTATTTGACCGGTCTATCC 1485
QY 1379 AAGCAATGGGACAGCCATTAAGCGGAGCCACCGAAGTGAATGAGCTGAGGATTCAT 1438
DB 1486 GAAACGAGCAGGAAGCTGGCGAGTCTCAGACGAGGTTGCTTCCAAAGTCCAAAGCGAAG 1545
QY 1439 GAGAAAGAGAGCGGTGACGACAGCGCTGGCAAGCTGGAAGCATGCTACTAAGCATGATCC 1498
DB 1546 AAGATGGCGGTGCGCAACGCCAATAAGAGCGGTCTCTTTTACAGGATTTTGAAGCGC 1605
QY 1499 CAGCAGGATATGCGGCTCAGCTCAACCTGAAGATGTTGTTATTTTGGCGGTTCTAATG 1558
DB 1606 AAGTCTGCTCGGATCTTTGGTCTCGACACTCGATTTGCTGTGTTTCCATCTCTCT 1665
QY 1559 CTGTTGATGATGTTTCGCTGTC---CACTGCACTGGGTGACCAAGCATGCTACTCCAGT 1615
DB 1666 GTCTTCTCTTCACTTTTGTCTTCTCACTGACATATGAGCTTCAACAGGATTTCTTGG 1725
QY 1616 CCCTCATGCTTGGCTTTCCACAGTCAGATGGATCCGCAACATTTTAGACGAT 1675
DB 1726 CTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1782
QY 1676 TTCAGAGAGGCTTACTGCTGCTTTCGCAAGCACTGCCGATGATGCTGCTGCTGCTGCTGCT 1735
DB 1783 TTCAGAGAGGCTTACTGCTGCTTTCGCAAGCACTGCCGATGATGCTGCTGCTGCTGCTGCT 1842
QY 1736 TGGTGGATTAAGGATACAGATAGCGGATAGCGCAACAGACAGCGCTAGTGGATAT 1795
DB 1843 TGGTGGATTAAGGATACAGATAGCGGATAGCGGATAGCGGATAGCGGATAGCGGATAGCGG 1902
QY 1796 AATAGCTGGAACATAGTCAATAGCGCTGTTGGCAAGGCAATGCTTCAACCGGAGGAG 1855
DB 1903 AATAGCTGGAACATAGTCAATAGCGCTGTTGGCAAGGCAATGCTTCAACCGGAGGAG 1962
QY 1856 AAGTCTAGCAATATGACATCTTGAAGTGAAGTACGCTGTTGCTGATCTTGGCGGT 1915
DB 1963 GTGCAATATCTTCTTGGAGAGCATGATGCTGATTTCTTGTGATCTTTGGGCGC 2022
QY 1916 GTGATGCGCTATTCGCGGATGATCAACAGATTCCTGTGGATGTTCCGAATTCGTGAG 1975
DB 2023 TTATTTGGGCTACTCTGTTGAGGATATCAACAGATTTTGTGGATGTTAGGATCTCACA 2082

QY 1976 GGAGAGCATCCCAAGGACATTAAAGGAAGCGATTACTTTACGACCGGGTGAATTCAGG 2035
DB 2083 GGTGAATGGCCTGACAGGTGAGGAGTCAACTACTTTACTCAAGAGGGGAGTATGCT 2142
QY 2036 GTAGATGCGGAAGGTGCTCGGCGCTTCAACTGCTTATGTACAAATTAAGTACTAC 2095
DB 2143 GTCGATGACAGGCGCACCCCTACTATGAAGAACTCTCTCATGTACAAATGCTTACTAC 2202
QY 2096 AGATTGCGGGAATTAAGTTGAGTACAGAGGTCCATCTGGATGATCGCACAGTAA 2155
DB 2203 CGCTTCCCGCA-----GCTTTATGTTGACACCGGCTCAAGACAGGTTGAGGC 2253
QY 2156 GCCTCATTTGGGAATAAGGACTTCGATCTGACTTACCTTGGAGGAGGCTTACACACAGAA 2215
DB 2254 CAAATATCCCTCTAAACAGTGTACTCTTGATATCTCTTGAAGAGGTTTCAATCCGAA 2313
QY 2216 CACTGCTTGTTCGATCTATAGGTTGAAGAG 2248
DB 2314 AATTGATGCTCAGGATCTTCAAGGTCAAGAG 2346

RESULT 15
US-10-425-115-150745
; Sequence 150745, Application US/10425115
; Publication No. US2004021427A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 150745
; LENGTH: 2839
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MR14577_59000C.1
US-10-425-115-150745

Query Match 25.6%; Score 619; DB 18; Length 2839;
Best Local Similarity 58.5%; Pred. No. 1.8e-179;
Matches 1244; Conservative 0; Mismatches 820; Indels 63; Gaps 7;
QY 121 CCTAATCACTTCGCCATCTCTGCTAATCGCTGGCTGGCGGATTTCTCTCGCCTCTT 180
DB 283 CTTGCTCCGGCTCTCGGCTGGCGCTCATCTACGTCTCGCTTCCGCTCGCCTCTT 342
QY 181 CGCGCTCATCGGTTTGAATTCGATTCATGAGTTGATCCGTTTCAACTACCGGCG 240
DB 343 CTCCGCTCTCGCTACGAGTCCATGATCCAGGTTCCAGCTTCACTTCACTACCGGAC 402
QY 241 CACCGCTACATGCTGTCAGAAATGTTGTTGATCAACTTCTCACTGTTTCAAGCGCGC 300
DB 403 CAGCTCTTCTTCAACGAGACGGCTTCAAGGATTTGAGAACTGTTTCACTTCACTTCAAG 462
QY 301 ATGATATCCGCTCGGAGGATTTGCGGCGTACCGCTTATCCCGCTGATGATGATGCTC 360
DB 463 CTGGTACCGCTCGGCGGCTTGTGCGGCGGAGCGCTTACCCGCGTCTGATGTTGAGCGC 522
QY 361 CGCGGGAATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 523 CGCGCTGCTTCCACCGCTCTCTCGGCGCTCTACTCTACCGCTTCACTTCACTTCAAG 582
QY 421 CGTGTTCGTCGCGCGGATTTTCAAGTGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
DB 583 CGTCTTCAACCGCGCGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 642
QY 481 GGAGCTGTGCTCGCGGCGCGCGCTCTTTCGCGCGGCTTCTTTCGCGCGGCTTCTTTCGCG 540



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 00:24:44 ; Search time 7349.2 Seconds
(without alignments)
11984.275 Million cell updates/sec

Title: US-10-028-384-7

Perfect score: 2417
Sequence: 1 tctaagcaagaatgtgtcg.....ccaaaaa..... 2417

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	960.6	39.7	2709	3	AK018758	Mus muscu
2	945.6	39.1	4513	3	BC044321	Xenopus 1
3	944	39.1	2669	3	AK012153	Mus muscu
4	817.2	33.8	836	9	CNS0172X	AL108711 Drosophil
5	807.4	33.4	810	2	BF502026	AT17657.5
6	711.8	29.4	715	7	CK657451	LP23759.5
7	710.8	29.4	736	1	AA949890	LD29946.5
8	699.4	28.9	701	2	BF503420	AT19406.5
9	688.4	28.5	690	4	BG640860	SD12024.5
10	681	28.2	681	7	CK659033	LP16462.5
11	680.4	28.2	694	4	BI354111	GM26338.5
12	677.4	28.0	679	7	CK657649	LP24119.5
13	676.2	28.0	682	7	CK659064	LP16523.5
14	659.8	27.3	663	4	BG641172	SD12448.5
15	658	27.2	666	4	BI357074	RE43425.5
16	658	27.2	668	4	BI227902	RE25288.5
17	653.4	27.0	655	2	BF496296	AT10060.5
18	652	27.0	652	4	BI171940	RE13841.5
19	648.4	26.8	655	4	BI484774	RE57926.5
20	644.4	26.7	658	4	BI374189	RE61718.5
21	644	26.6	644	4	BG636414	SD41423.5
22	644	26.6	654	4	BI374334	RE61893.5
23	641	26.5	641	1	AI257750	LP06212.5
24	635	26.3	642	4	BI173041	RE15774.5

25	633.4	26.2	635	1	AI135629	
26	631.2	26.1	3802	3	AK030363	Mus muscu
27	630	26.1	646	4	BI170481	RE11944.5
28	628.8	26.0	2656	3	BC048897	Mus muscu
29	628.4	26.0	630	1	AI295381	LP08987.5
30	625.6	25.9	649	4	BG641064	SD12296.5
31	623.6	25.8	643	4	BI162865	RR02124.5
32	621.8	25.7	2730	3	AK081547	Mus muscu
33	620.8	25.7	3643	3	AK077877	Mus muscu
34	614.4	25.4	616	2	BF504378	AT05729.5
35	612	25.3	615	4	BI364613	RE49773.5
36	603.2	25.0	2118	9	AY418284	Homio sapi
37	600	24.8	600	4	BI163592	RR03026.5
38	596.4	24.7	598	4	BI170396	RE11825.5
39	585	24.2	585	1	AI133998	GH11327.5
40	580.6	24.0	2727	3	AK087470	Mus muscu
41	576	23.8	585	4	BI242003	RE39217.5
42	568.6	23.5	590	1	AA694951	GM01838.5
43	568.2	23.5	2103	9	AY418286	Mus muscu
44	562.4	23.3	564	1	AI542327	SD08615.5
45	562	23.3	564	4	BI635692	SD17119.5

ALIGNMENTS

RESULT 1	AK018758	2709 bp	mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300006C19 product:hypothetical Oligosaccharyl transferase (Orafase) STT3 subunit containing protein, full insert sequence.	HTC 03-APR-2004
LOCUS	AK018758			
DEFINITION	AK018758.2	GI:26384577		
ACCESSION	AK018758			
VERSION	AK018758.2			
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	Carninci, P. and Hayashizaki, Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)			
MEDLINE	99279253			
PUBMED	10349636			
REFERENCE				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)			
MEDLINE	20499374			
PUBMED	11042159			
REFERENCE				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Iehikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, N., Inoue, Y., Kira, A. and Hayashizaki, Y.			
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multichannel capillary sequencer			
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)			
MEDLINE	20530913			
PUBMED	11076861			
REFERENCE				
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the PANTON Consortium.			
TITLE	Functional annotation of a full-length mouse cDNA collection			
JOURNAL	Nature 409, 685-690 (2001)			
REFERENCE				

HPKDIREDGYETQOGEPRVDKAGSPTLLNCLIMKWSYVRFGEMLDRTIPGPFORTN
AENKNDIKFKHLEBAFTSEHMLVRIYKVKAPDRETLGHKPRVTNIVPKQYLSKKT
TKRKGYYKNKLWFKKXKTSKIV
2893..2698
/note="putative"
2709
/note="putative"

polya_signal
polya_site

ORIGIN

Query Match 39.7%; Score 960.6; DB 3; Length 2709;
Best Local Similarity 64.8%; Pred. No. 6.2e-250;
Matches 1495; Conservative 0; Mismatches 794; Indels 18; Gaps 4;

QY 102 AGTGCTGGCTACAGCAGCCTAAACACCTTCGCCCATCTCTGTAATCGCCTGGCTGGCGG 161
Db 257 AGCGGCGCGGTGGCAGTGGTCTCTCTTCCATCATCTCTCTCTCTGGCTGGCTGGCGG 316
QY 162 GATTTTCCTCTCGCCTCTTTCGCCGTATCGCTTCGAGTCGATATTCATGAGTTTGATC 221
Db 317 GCTTCAGCTCGGCGCTCTTTCGCCGTATCGCTTCGAGTCGATATTCATGAGTTTGATC 376
QY 222 CGTGGTTCACTACCGGGCCACCGCTCATCGTGCAGATGCTTGGTACAACTTCTCTCA 281
Db 377 CGTGGTTCACTACCGGGCCACCGCTCATCGTGCAGATGCTTGGTACAACTTCTCTCA 436
QY 282 ACTGGTTTCAGCAGCGCGCATGATTCGCTTCGCCGAGATTCGCGCGGTACCGCTATC 341
Db 437 ATTGGTTTCAGCAGCGCGCATGATTCGCTTCGCCGAGATTCGCGCGGTACCGCTATC 496
QY 342 CCGGCTGATGATAGCTCGCGGGGATTCATTCGCTGCTGCGCTGCTGCTGCTGCTGCTG 401
Db 497 CAGGTTGATGATAAAGAGCTGGCTTATTCATTTGGAATTTAAATACATTGAACATAACG 556
QY 402 TCCATATTCGTGACATCTGCTTCTCGCGCGCATCTTCAGTGGCGCTGAGCTCCATCT 461
Db 557 TTCAATAGAGATGCTGCTTCTTCGCGCGCATCTTCAGTGGCGCTGAGCTCCATCT 616
QY 462 CCACCTACCTGCTGACCAAGAGCTGCTGCGCGCGCGCGCGCTCTTCGCGCGCGCTG 521
Db 617 CTAGTTTCTGCTAACTAGAGAACTGTGAAACCAAGAGCAGGACTTCTAGCTGCTGCT 676
QY 522 TCATCGCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 581
Db 677 TCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 736
QY 582 GCATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 841
Db 737 GCATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 796
QY 642 GATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 701
Db 797 GGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 856
QY 702 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 761
Db 857 GAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 916
QY 762 GCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 821
Db 917 AGAGGTACAGCAAGAGAGTCTACATGATATGACATTTGTTGATTTGCTGCTGCTGCTG 976
QY 822 TCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 881
Db 977 TATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1036
QY 882 CGCTGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 941
Db 1037 CTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1096
QY 942 TGTGCGCAACGAGTTCGGAGAGTGTTCATGCTGCGCGGATGCTGCTGCTGCTGCTGCTG 1001
Db 1097 TGACAAACAGAGGTTCAGAGACCTTTTCTTTTGGGTGCTGCTGCTGCTGCTGCTGCTG 1156

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation

JOURNAL of 60,770 full-length cDNAs

REFERENCE Nature 420, 563-573 (2002)

AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirose, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Okazaki, Y., Okido, I., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission

JOURNAL Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT On Dec 10, 2002 this sequence version replaced gi:12858635. Please visit our web site (http://genome.gsc.riken.jp/) for further details.

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCACTCGATGTTTCTTTTCTTNN 3'], cDNA was prepared by using triose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGTCCAGAGCTCAATATTTATTAACCGCCCGCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

Location/Qualifiers

1..2709

/organism="Mus musculus"

/mol_type="cDNA"

/strain="C57BL/6J"

/db_xref="FANTOM_DB:1300006C19"

/db_xref="taxon:10090"

/clone="1300006C19"

/sex="male"

/tissue_type="liver"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

76..2547

/note="unnamed protein product; hypothetical oligonucleotide transferase (OTase) SRT3 subunit containing protein (InterPro|IPR003674, evidence: InterPro)

putative"

/codon_start=1

/protein_id="BAB311390.1"

/db_xref="GI:12858636"

/translation="MAPSPAPESKHSSLNSSPWSGLMALGNRSHGHGPGTQASRA AAPKPGAGLGGSLGSPAGWQSLLSFTLLFLAWLAGPSRLFAVIRFESIRHEDPW FNVSRHLHSHGFYELNWDERAYPLRGVYVPGMLITAGLHILHILNLTNI VHRIDVCVFLAPFSGTSLSTELLRELWNOGAGLLAIFAIVEGYISRVAGSFD NEGIAFALOPTVYLVKSVKTSVFTWCCCLSYFVMSWAGYVFIINLPLHVFV LLLNORYSKVEFTLVYLVGLILNQIPFVGFOPRTSEHMAAGVALLQAVFL QYLRDLTKREFQTLPLGVSLLAGAVLSVILYTVGYIAPWSGREYSLDWTGAKI HIPLIADSVSEHPTVWVSFFDFDLHILVCTFPAGLWFCIKINDEKRVFVYALISAVF AGVMVRLMLTLPVVCMLSAIFSNVFEHYLDGDMKRENPDESDEDDKKNPNLY DKAGVRKHVTEQKPEGLGPNIKSIVTLMMLLMLEFAVHTCTWTSNAYSPPSVL ASYNHGDGRNLDLDFRAYFWLRQNTDEHARVMSWMDYGYIAGMANRFTLVNDNTW NSHTALVGVKAMSSNETAYKIMRSLDVYVYLVIFGGVIGYSGDDINKFLMVRABE

QY 1002 TCTTTGTCGGCGTGTGCTGCTCACCATGCTGGGGTGTGGCCCGTGGAGTGGACGCT 1061
 DB 1157 TGTTCCCTTAGTGTCTATCTATCTGACATACACAGGTATATTTGCAACCATGGAGTGGCAGGT 1216
 QY 1062 TCTACTCGCTCTGGGATACCTGGTACGCGCAAGATCCCAATTCCTCATTTGCACTCCGCTGT 1121
 DB 1217 TTTTATCTACTATGGGATCTGGGTATGCAAAATATACCAATTCCTCATTTGCACTAGGT 1276
 QY 1122 CGGAGCATACGCCACCACTTTGGTCTCGTTCTTTCTTTGATCTGCAATCTCTGGTGGCG 1181
 DB 1277 CTGAACATACGCTACGACATCGGTGTCTTTCTTTCTTTGATCTACATATTTCTTTGATGTA 1336
 QY 1182 CTTTCCCAAGTGGGATGTGGTACTGCAATCAAGACAGATCAACGAGCGGGTCTTTCTGGTG 1241
 DB 1337 CTTTCCCAAGCGGCTATGTTCTGCAATCAAAATATCAAGATGAAGAGTATTTGTCG 1396
 QY 1242 TGCTGTAGCCCATCAGTGGGTTTACTTCTGGTGTGTGATGGTGGTTCATGTTGACCC 1301
 DB 1397 CTCTGTATGCGATCAGTGTCTGTACTTTGCGGAGTGTGGTGGGCTGATGCTGACTC 1456
 QY 1302 TCAGCGCGGTGTGTGATGCTGCGGAGTGGGCTTTTTCGGGACTGTTGGATGTTGTC 1361
 DB 1457 TGACCCCGGTGTGTGATGCTGCGGATGCGCTTCTCAATGTTTTTGAGCACTATT 1516
 QY 1362 TGAAGAGGATTCGTCTAAGCGAATGGGCAAGCCATTAAGCGAGCCACCGAAGTGGATG 1421
 DB 1517 TGGGGGATGA-----CATGAAAGGGAACCCACCTGTGGAG-----GACAGCAGTG 1564
 QY 1422 AAGCTGAGGATTCATTTAGAGAGAGCGCTGTACGACAGGCTGGGCAAGCTGAAGATC 1481
 DB 1565 ATGAGGATGACAAAGAAACCCAGGAACCTTTGATGACAGGCGGTAAAGTGGAGAAC 1624
 QY 1482 GTACTAAGCATGATGCCAGCAGGATATCTGGCGTCAAGTCCAACTCAAGAGTATTTGTTA 1541
 DB 1625 ATGTGACAGAGCAAGAGAAACCTGAAGAGGCTTGGSCCCCAACATCAAAAGCATTTGTA 1684
 QY 1542 TTTTGGCGGTTCAATGCTGTGATGATGTTGCTGTCTCACTGAGTGGGTGACCAAGCA 1601
 DB 1685 CCATGTGATGCTATGCTCTGATGATGTTGCGGTCCACTGCGTGGGTGATGATGATG 1744
 QY 1602 ATGCTACTCCAGTCCCTCCATTTCTTGTGCTTTCACACAGTCAAGATGGATCCGCA 1661
 DB 1745 ACGCTACTCCAGTCCAGTGTGCTCTGCTCTCTCAAA---TCATGATGGTACCAGGA 1801
 QY 1662 ACATTTAGAGATTTCAAGAGCTTACTTGTGCTTTCGAGAAACACTGCGATGATG 1721
 DB 1802 ATATATTAGATGATTTTAGAAGAGCGTACTTTTGGCTGAGACAAACACCGATGAACG 1861
 QY 1722 CTCGCTTATGCTCTGTGGGATTTAGGATACAGATAGCGGAATGGCAACAGAACCA 1781
 DB 1862 CCGGGTCAATGCTGTGGGATCTACGGCTATCAGATTTGTCGATGGGCAACAGGACCA 1921
 QY 1782 CGCTAGTGGATAATACTGGAACAATAAGTACATAGCGCTGTGGTGGCAAGGCAATGT 1841
 DB 1922 CTCTGTGGGATAACAACCCCTGGAACAACAGCCACATCGCACTGGTGGAAAGCTATGT 1981
 QY 1842 CTTCAACCGAGGAGAGTCTAGCAATATGATCATCTCTGTGAGTGGACTAGTGGTGG 1901
 DB 1982 CTTCCAAATGAACGGCGCTATAAAATCATGAGTCCCTGTGATGTCGATGATGTTGG 2041
 QY 1902 TGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGATATCAACAGTTCCTGTGGATGG 1961
 DB 2042 TTATTTTCGAGGAGTATGGCTATTTCGGGAGCGATATCAACAGTTCCTCTGGATGG 2101
 QY 1962 TCCGAATGCTGAGGAGAGCATCCCAAGGACATTAAGGAAGCGATTAATTACCGGACC 2021
 DB 2102 TCAGGATAGCTGAAGGGAGCATCCCAAGACATCCGGGAAGGTGACTATTTCAACCCAGC 2161
 QY 2022 CGCGTGAATTCAGGTAGATGCCGAAGGTGCTCCGSCCTGTCTCACTGCTCTATGTTGTA 2081
 DB 2162 AGGGAGGTTCAGGTAGACAAAGTGGTCTCTCTACTCTGTTAACTGCTCTATGTTATA 2221
 QY 2082 AATTAAGCTACTACAGATTCCGGGAATTAAGTTGGACTACAGAGTCCATCTGGATATG 2141

DB 2222 AAATGTCTACTACAGATTGGAGAAATGCAGTAGATTTTCGCACTCCCCAGGCTTTG 2281
 QY 2142 ATGCGACACGTAAACCGCTATTGGGATAAAGGACTTCGATGTAACCTGACCTGGAGGAG 2201
 DB 2282 ACGAAACAGTATGCTGTGATTTGGAATTAAGACATTAATTAAGCATTTGGAGGAAG 2341
 QY 2202 CTTACACCAAGAACACATCGGCTTTGTCATCTATAGGTGAAGAGCGCATGATGTTCA 2261
 DB 2342 CTTTATCATCAGAGCACTGGCTTGTGAGGATATAAAGTGAAGACACCTGACACAGGG 2401
 QY 2262 ATAGACCATCACTGAGACCAAGGAGAGAACGATTCCTCCAGCAACTTCATTTCCAGAA 2321
 DB 2402 AGACCTAGGTCAAAACCTCGAGTCAACCAATCGTCCCAAAACAGAAATTTGTCAA 2461
 QY 2322 AGAAGCTTAAGCGTGGCAAG---GGCTACATACAGAAACCGACCGGTGTTGTTAAGGAA 2378
 DB 2462 AGAAGACTACTAAAGGAGCGTGGCTAGTTAAATTAAGCTAGTGTGTTAAGAAAGCA 2521
 QY 2379 AACGACCTTGAATTAACCCCAAAAAA 2405
 DB 2522 AGAAGACCTCTAAGAAGACTGTTTAAA 2548

RESULT 2

BC044321 4513 bp mRNA linear HTC 23-JAN-2003
 LOCUS
 DEFINITION
 XENOPUS LAEVIS, clone IMAGE:4684177, mRNA.
 BC044321
 VERSION
 BC044321.1 GI:27882209
 KEYWORDS
 HTC.
 SOURCE
 XENOPUS LAEVIS (African clawed frog)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopus laevis; Xenopus.
 REFERENCE
 1 (bases 1 to 4513)
 AUTHORS
 Klein, S. and Strausberg, R.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (17-JAN-2003) National Institutes of Health, Xenopus Gene
 Collection (XGC), National Institute of Child Health and Human
 Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
 20892-7510, USA
 REMARK
 NIH-MGC Project
 Contact: XGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. Igor Dawid
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 CDNA Library Arrayed by: Institute for Systems Biology
 http://www.systemsbio.org
 contact: amadan@systemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 94 Row: i Column: 19
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, similarity but not identity to protein
 This clone has the following problem: frame shifted.

FEATURES

source
 1..4513
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:4684177"
 /tissue_type="Embryo, stage 31/32, Xenopus"
 /clone_lib="NICHHD_XGC_Emb4"
 /lab_host="DH10B"
 /note="vector: pCMV-SPORT6"

ORIGIN

Thu Dec 16 16:25:14 2004

Query Match	39.1%;	Score	945.6;	DB	3;	Length	4513;	
Best Local Similarity	64.4%;	Pred. No.	8.5e-246;					
Matches 1491;	Conservative	0;	Mismatches	799;	Indels	26;	Gaps	4;
QY	102	AGTGGCTGGCTACAGCAGCTTAATCAGCTTCGGCATCTGCTTAATCGCTGGCTGGCGCG	161	QY	1122	CGGAGCATCAGCCACCACTTGGTTCTCGTTCTCTTTTGTGATCTGCAATCTCGTGTGGC	1181	
DB	446	AGCCGGGGCTGGCAGTCTCTGCTCTCTTCACTATTCATTCCTGGCTGGCTGGCG	505	DB	1466	CTGAACATCAACCAACATGGGTGCTCTCTTTTGTGATCTGATATTTAGTATGTA	1525	
QY	162	GATTTTCCTCTCGCTCTTCGGCGTATCCGTTTCAGTTCGATTCATTCATGAGTTTCATC	221	QY	1182	CTTCCAGTGGAGTGTGTTACTGATCAACAGCATCAACAGAGAGCGGCTTTTCGTGG	1241	
DB	506	GCTTCAGCTCAAGGCTCTTCGCTGTATCCGCTTCGAGAGCATCATCCAGAGTTCGACC	565	DB	1526	CATTTCCGTGAGGACTTTGGTTCTGCATAAAATATCAACGATGAAGAGTTTGTG	1585	
QY	222	CGTGGTTCACACTACCGGCCACCGCTACATGTTGTCAGAAATGTTGTTACAACTTCCTCA	281	QY	1242	TGCTGTACCCCATCAGTGGCTTTACTTCGTGTGTGATGGTGGTTTGTGATGTGACCC	1301	
DB	566	CTTGTTTAACTACAGATCTACATCAGCTAGCATCCATCCATGGCTTTTATGAATTTTAA	625	DB	1586	CTCTGTATGCAATTAGTCCGTTTACTTTGTGGCGTAAATGGTTGTTGATGTGACCT	1645	
QY	282	ACTGTTTCAGAGCGCGCATGTTATCCGCTCGGAGGATTTGGGGCGGTACCGTCTATC	341	QY	1302	TCAGCCGGTGTGTGATGCTGGCGGAGTGGCTTTTCGGGACTGTTGGAATGTGTTCC	1361	
DB	626	ATTGGTTTGAACCGGCATGTTATCCATCTGGGAGAAATAGTAGTGGAAACAGATATACC	685	DB	1646	TGACTCCTGTGTTTGCATGCTGGCCATTCGCTCTCTCAATGTTTTCGAGCATATC	1705	
QY	342	CGGCTGTGATGATTAAGCTCGCGGGAATCCATTGGCTGTGCAAGTACTCAACATACCG	401	QY	1362	TGCAAGAGGATTCGTTCTAAGCGAATGGGCACAGCCATTAAGCGCAGCCACCGAAGTGGATG	1421	
DB	686	CTGGATTAATGTTTACAGCTGGGCTCATCCATTTGGAATTTTAAACATGTTTAAACATACCG	745	DB	1706	TGGGTGATGATATGAAGCGGGAACCCGCCAGTGGAGACAGCAGTGCAGAGAGCGAA	1765	
QY	402	TCCATATTCGTGACATCTGGTGTTCCTGGCGCCGATCTTCAGTGGCTGACCTCCATCT	461	QY	1422	AAGCTGAGGATTCATTTGAGAAAGACGCTGTACGACAGGCTGGCAAGCTGAGCATC	1481	
DB	746	TTTATATACAGATGTATGTGTCTTTTAGCACCATGATTTAGTGGCTTACATCAATCT	805	DB	1766	ATTCTG-----GTACACTCTATGATAAGGCTGGCAAGTGAAGAGC	1807	
QY	462	CCACTACCTGTGACCAAGAGCTGTGTCCGCGGCGCGCTCTTCGCGCGCAGCT	521	QY	1482	GTACTAAGCATGATGCCAGCAGGATCTGGGCTCAGCTCCCAACCTGAAGATATTGTTA	1541	
DB	806	CCACTTCTGCTCACTCGAGAACTGTGAAACAGGAGCGGGAATCTTAGCTGCTGCT	865	DB	1808	ATGTGTCGAGAGGAAAGCAGAGGAGGCTTTCGCCCTAATATAAAGAGTATAGTCA	1867	
QY	522	TCAATGCCATGTCCTGGCTTACATCAGTAGTGTGGTGGCTGATCGTACGATACAGG	581	QY	1542	TTTTGGCCGTTCTAATGCTTGTGATGATTTCCGTGTCCACTGCACGTGGGTGACGCA	1601	
DB	866	TCATTCGTATAGTCCAGGATACATATCCAGATCAGTAGCTGATCATTTGACATGAAG	925	DB	1868	CTATGTTGATGCTAATGCTGTTGATGATGTTTGTGTACACTGTAGCTGGGTAAC	1927	
QY	582	GCAATGCGATTTTCGCGCTGAGTTCACCTACTTCTGTGGGTGGCTCAGTGAAGCTG	641	QY	1602	ATGCTTACTCCAGTCCCTCCATGCTTGGCTTTCCACAACTCAAGATGATCCCGCA	1661	
DB	926	GCAATGCTATCTTTGCACTGCAATTTACATATATTTTGGGTAATAATCTGTAAACAG	985	DB	1928	ATGCTTACTCCAGCCCTAGTGTGTTATTTGGCTCATCAA---TCAGNTGAAACAGAA	1984	
QY	642	GATCCGTTTCTGTCGGCGCAGCGCTTGTCTACTTCTCATGTTGTGCTGGCTGGG	701	QY	1662	ACATTTTAGACGATTTTCAGAGAGGCTTACTACTGCTTTTCGAGAACACTGCGGATGATG	1721	
DB	986	GCTCTGTTCTGCGCAATAGTTGTGTCTGTCTTATTTTATATGTTTTCGCGCTGGG	1045	DB	1985	ATATCTTGATGATTTTCAGAGAGCATACTACTGGCTAAGGAGAGATACAGATGAACATG	2044	
QY	702	GTGGCTACGTTTTCATCATCAACTGATACCCCTGCACTCTTGTGTAATGCTCATATGG	761	QY	1722	CTCGGTTATGCTTGGTGGGATTACGGATACCGATAGCGGGAATGGCAACAGAACCA	1781	
DB	1046	GTGGTTATGTTTATCAATTAATCTTATTCACATACATGTTTGTACTGTTACTGATGC	1105	DB	2045	CCCGTGTATGCTCCTGGTGGGATTATGGCTTACCAAAATTCAGAAATGGCTTATCGAACCA	2104	
QY	762	GCAGGTACTCGCCGCTGCTGTGACAGCTACAGCACCCTTACATCTCGGACTGCTGT	821	QY	1782	CGCTAGTGGATAATAATACGTGGAAACAATAGTACATAGCGCTGCTGGTGGCAAGCAATGT	1841	
DB	1106	AGAGATACAGCAAGAGTCTACATAGCTTACAGCACTTTTACATTTGGGTTTAAATAC	1165	DB	2105	CTTTAGTAGACAACTACGTGGAAACAATAGCCACATAGCTTTGGTTGGAAAGCAATGT	2164	
QY	822	TCTCCATGCAATCCCTTCGTGGGATTCACCGGATACGCAACAGTGAACACATGGCTG	881	QY	1842	CTTCAACCGGAGGAAAGTCTTACGAAATTAAGCAATCTCTTGACGTGGACTAGCTTTGG	1901	
DB	1166	TATCAATGCAATTCCTTTGTGGGATTCAGCCAAATACGCAAGTGAACATATGGCAG	1225	DB	2165	CTTCAAAATGAACCGCTGCTTATGAATAATGAAGACTTGGATGTAGATTTATGTAATA	2224	
QY	882	CGCTGGGAGTGTGTGCTCTTATGGCGTGGCCACCTTGGCCATTTGAGTCCGCTGC	941	QY	1902	TGATCTTTTGGCGGTGATCGGCTATTCTGGCGATGATATCAACAAGTTCTCTGTGGATGG	1961	
DB	1226	CTGAGGGGTCTTTGCCCTGCTGCAAGCTTATGCTTCTTCTGCTGATGCTGGAGCTG	1285	DB	2225	TAAATTTTGGGGTGTAAATTTGGATATTCGTGTGATGACATCAACAATACTCTCGATGG	2284	
QY	942	TGTCCGCCAACAGATTCGGGAAGCTGTTCATCGTCGGCGGATTTGCTGGGGCTTTGGC	1001	QY	1962	TCCGAATTTGCTGAGGGAGAGCATCCCAAGGACATTAAGAAACCGATTTACTTTACCGACC	2021	
DB	1286	TAAGTAGCAGGATTTCCAGACACTGTTTTCCTCGGTGTATCGTTGGCAGCTGGAGCTG	1345	DB	2285	TTCCGAATAGCAGAGGAGGAACATCCAAAGATATCAGGGAAAGTGAAGTCTACTTCCACC--AC	2342	
QY	1002	TCCTTGTGGCGCTGTGGTGTCTACCATGCTGGGGGTTTGGCCCGCTGGAGTGGAGCT	1061	QY	2022	GGGTGTAATTCAGGTTAGATGCCAGGTGCTCCGGCCCTGCTCAACTGCTTATGTACA	2081	
DB	1346	TATTTCTCACTGTCACTATTTGACTTATACAGGCTTACATGCTTCCATGAGTGGAAAGT	1405	DB	2343	AAGGAGAGTTTCTGTGTAGACAAAGCTGGATCTCCAACTTGTCTCAATTCGCTCATGTATA	2402	
QY	1062	TCTACTCGCTGTGGGATCTGGCTACGCCCAAGATCCACATTTCCCATTCATTCGATCCGCTGT	1121	QY	2082	AATTAAGCTACTACAGATTCGGGGAATTTGAAGTTTGAAGTGTGACTACAGAGGTCCATCTGGATATG	2141	
DB	1406	TTTACTCACTTTGGGACACAGGGTATGCGAAATCCATATCCCATTCATTCGATCTGTGT	1465	DB	2403	AGATGTCAATTTACCGGTTTGGTGAATGCAGCTTGAATTTCCGAACACCACTGCGATTTG	2462	


```

Db      2523 CATTACATCGGACCTGGCTGTGTTAGAAATATACAAAGTTAAAGAGTCAGAAACAGAG 2582
Qy      2262 ATAGACCATCATCTGAAGACCAAGGAGAGAAAGTTCCTCCAGCA---AACTTCATTTCGA 2318
Db      2583 AGGCATTGATCATAAACCTAGAGTAACCAATATTGTACCAAAACAGAAATATTATCAA 2642
Qy      2319 GAAAGAACTCTAAGCGTCCGACGGGCTCATACGAAACCGACCGTGTGTTGTTAAGGAA 2378
Db      2643 AAAGACATCTAAGAGGAAGCGTGCTACATTAGGAACAAGCTGACACTGAAGAAAGCA 2702
Qy      2379 AACGAACTTGAATAAACCCMAAAAAA 2414
Db      2703 AGAGCAACAAAGAAGTCTGTCTAAACAACTTAA 2738

RESULT 3
LOCUS   AK012153                2669 bp      mRNA      linear      HTC 03-APR-2004
DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
Oligosaccharyl transferase (OTase) STT3 subunit containing protein,
full insert sequence.
ACCESSION AK012153.1 GI:12848721
VERSION   1
KEYWORDS HTC; CAP trapper.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1
AUTHORS   Carninci, P. and Hayashizaki, Y.
TITLE     High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   93279253
PUBMED    10349636
REFERENCE 2
AUTHORS   Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED    11042159
REFERENCE 3
AUTHORS   Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED    11076861
REFERENCE 4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium
TITLE     Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS   The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE     Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS   Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,

```

Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAATAATCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.

Host: SOLR.

FEATURES

source

Location/Qualifiers

1..2669
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:2610524N02"
/db_xref="taxon:10090"
/clone="2610524N02"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days embryo"

misc_feature

36..2507
/note="hypothetical Oligosaccharyl transferase (OTase) STT3 subunit containing protein (InterPro|IPR003674, evidence: InterPro)
putative"

polyA_signal

2654..2659
/note="putative"

polyA_site

2669
/note="putative"

ORIGIN

Query Match 39,1%; Score 944; DB 3; Length 2669;
Best Local Similarity 64.9%; Pred. No. 2.1e-245;
Matches 1498; Conservative 0; Mismatches 790; Indels 20; Gaps 6;
Qy 102 AGGTGGCTGGTCAGACAGCCTAATCATTGCGCATCTCTGCTAATCGCTGGTGGCGCG 161
Db 217 AGCGGCGGGTGGCAGTCGTGTCTCTCTTCCATCTCCTTCCATCTCTCTGCGCTGGCGCG 276
Qy 162 GATTTTCTCTCGGCTCTTCGCGTCATCGGTTTCGAGTCGATATCCATGAGTTTGATC 221
Db 277 GCTTCAGCTCGGCTCTTCGCGTCATCGGTCATCGGTCGAGAGCATCATCCAGAGTTCGACC 336
Qy 222 CGTGGTTCAACTACCGGGCCACCGCTTACATGGTGCAGAAATGGTTGGTACACTTCTCA 281
Db 337 CGTGGTTTAACTATAGATCAACATCATCTCTCATCTCATGATCTATGAGTTTCTAA 396
Qy 282 ACTGTTTCGACGCGCGCATGTATCCCGTCGCGAGGANTTGGCGGACCGTATC 341
Db 397 ATTGGTTTGAAGAAAGCAGCATGTATCCCACTGGGAAAGATAGTGGTGGCACCCTTTACC 456
Qy 342 CCGGCTCTGATGATTAGTTCGCGGGGAATCCATTGGTGTGTCGACGTACTCAACATACCGG 401

Db 457 CAGGGTTGATGATAACAGCTGGCTTATTCATTGGAT-TTAAATACATTGAACATAACAG 515
 Qy 402 TCCATATTCGAGCATCTGGGTGTTCTGCGCGGATCTTCAGTGGCCTGACCTCCATCT 461
 Db 516 TTCATTAAGAGATGTGTGTATTTCTTGACCAACTTTTAGCGGCTTATCATCATAT 575
 Qy 462 CCACCTACCTGTGACCAAGAGCTGTGTGTCGCGGCGCGGCTTTCGCGCGAGCT 521
 Db 576 CTAGCTTCTGCTAACTAGAGAACTGTGGAACCAAGAGCAGACTCTTAGCTGCTTCT 635
 Qy 522 TCATCGCATCTGCGCTGCTACATCAGTAGGTGCTGCTGATGCTGATGCTGATGCTG 581
 Db 636 TCATGCTATCTGACAGAGTACATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 695
 Qy 582 GCATTCGATATTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 641
 Db 696 GCATTCGATATTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 755
 Qy 642 GATCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 701
 Db 756 GGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 815
 Qy 702 GTGCTACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761
 Db 816 GAGGTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 875
 Qy 762 GCAGTACTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 821
 Db 876 AGAGTACAGCAGAGAGTCTACATGATATAGCACTTCTACATTTGGGTATTAAT 935
 Qy 822 TCTCCATGAGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 881
 Db 936 TATCCATGAGATACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 995
 Qy 882 CGCTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 940
 Db 996 CTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
 Qy 941 CTGTCGGGCAACAGATTCGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1000
 Db 1056 TTGACAAAACAGAGTTCAGAGCCCTTTCTTTTGGGTGCTGCTGCTGCTGCTGCT 1115
 Qy 1001 GTCTTTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1060
 Db 1116 GTGTTCTTAGTGTCTATCTGACATACAGAGTATATGACCATGAGTGGCAG 1175
 Qy 1061 TTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1120
 Db 1176 TTTTATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
 Qy 1121 TCGAGATACAGCCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1180
 Db 1236 TCTGAAATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1295
 Qy 1181 GCCTTCCAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1240
 Db 1296 ACCTTCCAGAGCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
 Qy 1241 GTGCTGACGCTACAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1300
 Db 1356 GCTCTGTATGGATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
 Qy 1301 CTCAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1360
 Db 1416 CTGACCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1475
 Qy 1361 CTGCAAGAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1420
 Db 1476 TTGGGG-----GTGACATGAAGAGGAAACCCACCTGCTGGAG-----CACAGCAGT 1523
 Qy 1421 GAAGCTGAGGATTCATTTGAGAGAGACGCTGTGCTGACAGAGCTGCGAGCTGAGGAT 1480

Db 1524 GATGAGGATGACAAAGAAACCCAGGAAACTTGTATGACAGCAGGTAAGTGAAGAG 1583
 Qy 1481 CGTACTTAAGCATGATGCCAGCAGGATATCTGGCTCAGCTCAAACCTGAAGAGTATTGTT 1540
 Db 1584 CATGTGACAGCAAGAAACCTGAAGAGGCTTGGGCCCAACATCAAAAGCATTTGTG 1643
 Qy 1541 ATTTTGGCGCTTCTAATGCTGTGTATGATGTTGCTGCTCACTGACAGTGGTGACAGC 1600
 Db 1644 ACCATGCTGATCTCATGCTCTGATGATGTGCGGTGCTGCTGCTGCTGCTGCTGCT 1703
 Qy 1601 AATGCTTACTCCAGTCCCTCCATGCTTGTGCTTCCACACAGTCAAGATGATCCCGC 1560
 Db 1704 AAGCTTACTCCAGTCCAGTGTGCTTCTGCTTCTTACAA---TCAATGATGTTACCAG 1760
 Qy 1661 AATTTTATGACGATTTAGAGAGGCTTACTACTGCTTTCGAGAACACTGCGGATGAT 1720
 Db 1761 AATATATTAGATGATTTTAGAGAGCGTACTTTTGGCTGAGACAAACACCGATGAAC 1820
 Qy 1721 GCTCGGCTTATGCTTGTGCTGATTTACGATACAGATAGCGGAATGGCAAAACAGAAC 1780
 Db 1821 GCCCGGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1880
 Qy 1781 ACCTTATGCTGATTAATACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1840
 Db 1881 ACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1940
 Qy 1841 TCTTCAACCGAGGAGAGTCTTACGAAATATGACATCTTGTGCTGCTGCTGCTGCTGCT 1900
 Db 1941 TCTTCAATGAACCGCGCTTATTAATATCATGAGTCTTGTGCTGCTGCTGCTGCT 2000
 Qy 1901 GTGATCTTTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1960
 Db 2001 GTTATTTTCCGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2060
 Qy 1961 GTCCGAATTTCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTTACCG 2020
 Db 2061 GTCAGATAGCTGAAGGGGAGCATCCCAAGACATCCGGAAGGTGACTATTTTCAACC 2120
 Qy 2021 CGCGGTGATTCAGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2080
 Db 2121 CAGGAGAGTTCGAGTGAACAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2180
 Qy 2081 AATTAAGCTACTACAGATTTGGGGAATGAAGTGGACTACAGAGTCTCATCTGCTGAT 2140
 Db 2181 AAAATGCTACTACAGATTTGGGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2240
 Qy 2141 GATCGCACACGCTAAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2200
 Db 2241 GACCGAACACGTAATGCTGAGATTTGGAATTAAGACATTAATTAAGCATTTTGGAG 2300
 Qy 2201 GCTTACACACAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2260
 Db 2301 GCTTTTATCATCAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2360
 Qy 2261 AATAGACCATCTGAAAGCAAGAGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2320
 Db 2361 GAGACACTAGGTCTCAAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2420
 Qy 2321 AAGAACTCTAAGCGTGGCAAG---GGCTACATAGAAACCGACCGGTTGTTTGAAGG 2377
 Db 2421 AAGAGACTACTAAAGAGAGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2480
 Qy 2378 AAGAGACTTGAATTAACCCCAAAAA 2405
 Db 2481 AAGAGACTTGAAGAGCTGTTTAAA 2508

RESULT 4
 CNS017ZX
 LOCUS

836 bp DNA linear GSS 26-JUL-1999
 Drosophila melanogaster genome survey sequence T7 end of BAC
 BACNS7M18 of DrosBAC library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.

Margaret Fuller. Sized fractionated cDNAs were directly ligated into pOTB7. Plasmid cDNA library."

```

ORIGIN
Query Match      33.4%; Score 807.4; DB 2; Length 810;
Best Local Similarity 99.8%; Pred. No. 2.6e-208;
Matches 808; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 23 GCATTCAGATCGGTATTAATTTTCGAGTTACTGGCTGGAATTTGGACATGAATCGGAG 82
DB 1 GCATTCAGATCGGTATTAATTTTCGAGTTACTGGCTGGAATTTGGACATGAATCGGAG 60

QY 83 CGAAGATGCTGAACAGCAAGTGGCTGCTACAGCAGCCTATACCTTCGCCATCTTG 142
DB 61 CGAAGATGCTGAACAGCAAGTGGCTGCTACAGCAGCCTATACCTTCGCCATCTTG 120

QY 143 CTAATCGGCTGGCTGGCGGATTTTCCTTCGGCTCTTCGCCCTCATCGGTTTCGAGTCG 202
DB 121 CTAATCGGCTGGCTGGCGGATTTTCCTTCGGCTCTTCGCCCTCATCGGTTTCGAGTCG 180

QY 203 ATTATCCATGATTTGATTCGGTTCATACCTACCGGCGACCGCCTACATGTTGACAGAT 262
DB 181 ATTATCCATGATTTGATTCGGTTCATACCTACCGGCGACCGCCTACATGTTGACAGAT 240

QY 263 GGTGGTACAACTTCCTCAACTGGTTCGAGGCGGCGCATGTATCGGCTCGGCGAGGAT 322
DB 241 GGTGGTACAACTTCCTCAACTGGTTCGAGGCGGCGCATGTATCGGCTCGGCGAGGAT 300

QY 323 GTGGCGGTACGCTATCCGGGCTGATGATTTACGTCGGCGGGAATTCATTGGCTGCTG 382
DB 301 GTGGCGGTACGCTATCCGGGCTGATGATTTACGTCGGCGGGAATTCATTGGCTGCTG 360

QY 383 CACGTACTCAACATACCGGTCATATTCGTGACATCTCGCTGTTCGCGCGCATCTTC 442
DB 361 CACGTACTCAACATACCGGTCATATTCGTGACATCTCGCTGTTCGCGCGCATCTTC 420

QY 443 AGTGCGCTGACCTGATCTCACTACTGCTGACACGAGAGCTGTGCTCGGCGGCGC 502
DB 421 AGTGCGCTGACCTGATCTCACTACTGCTGACACGAGAGCTGTGCTCGGCGGCGC 480

QY 503 GGCCTCTTCGCGCGGCTTCATCGCCATCGTGTGCTGCTGCTGCTGCTGCTGCTGCT 562
DB 481 GGCCTCTTCGCGCGGCTTCATCGCCATCGTGTGCTGCTGCTGCTGCTGCTGCTGCT 540

QY 563 GGATGCTACGATACGAGGATTTGCCATATTTGCGCTGCTGCTGCTGCTGCTGCTGCT 622
DB 541 GGATGCTACGATACGAGGATTTGCCATATTTGCGCTGCTGCTGCTGCTGCTGCTGCT 600

QY 623 GTGCGCTCAGTGAAGACTGGATCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 682
DB 601 GTGCGCTCAGTGAAGACTGGATCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660

QY 683 TACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 742
DB 661 TACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720

QY 743 TTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 802
DB 721 TTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780

QY 803 TACATCTGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 832
DB 781 TACATCTGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 810

```

RESULT 6

CK657451

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

715 bp mRNA linear EST 30-JAN-2004
 LP23759 Sprime LP Drosophila melanogaster larval-early pupal pOT2
 Drosophila melanogaster cDNA clone LP23759 5, mRNA sequence.

CK657451

CK657451

CK657451.1

GI:41400976

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 715)

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S. and Rubin, G.M.

BDGP/HMM: Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

Plate: LP.237 row: E column: 11

High quality sequence stop: 644.

Location/Qualifiers

1..715

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP23759"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DHS-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal

pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI;

Site 2: XhoI; Sized fractionated cDNAs were directly

ligated into pOT2. Plasmid cDNA library.

ORIGIN

```

Query Match      29.4%; Score 711.8; DB 7; Length 715;
Best Local Similarity 99.7%; Pred. No. 2.8e-182;
Matches 713; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 AATGTCTGTTGCAATTCAGATCGGTATTAATTTTCGAGTTACTGGCTGGAATTTGGACA 71
DB 1 AATGTCTGTTGCAATTCAGATCGGTATTAATTTTCGAGTTACTGGCTGGAATTTGGACA 60

QY 72 TGAATCGGACGCGAGATGCTGAACAGCAAGTGGCTGCTACAGCAGCCTATACCT 131
DB 61 TGAATCGGACGCGAGATGCTGAACAGCAAGTGGCTGCTACAGCAGCCTATACCT 120

QY 132 TCGCCATCTGCTAAATCGCTGGCTGGCGGATTTTCCTTCGCTCTTCGCGCTATCC 191
DB 121 TCGCCATCTGCTAAATCGCTGGCTGGCGGATTTTCCTTCGCTCTTCGCGCTATCC 180

QY 192 GTTTCAGTCTGATTAATTCATGATTCGCTGCTCAACTACCGGCGCACCGCTACA 251
DB 181 GTTTCAGTCTGATTAATTCATGATTCGCTGCTCAACTACCGGCGCACCGCTACA 240

QY 252 TGGTCGAGAATGGTTGGTACAACTTCCTCACTGGTTGACAGAGCGCGCATGGTATCCGC 311
DB 241 TGGTCGAGAATGGTTGGTACAACTTCCTCACTGGTTGACAGAGCGCGCATGGTATCCGC 300

QY 312 TCGGCAGGATTTGGGCGGTACCGTCTATCCGGGCTGATGATTACGTCGGCGGGAATCC 371
DB 301 TCGGCAGGATTTGGGCGGTACCGTCTATCCGGGCTGATGATTACGTCGGCGGGAATCC 360

QY 372 ATTGGCTGTGACGCTACTCAACATACCGGTCCTATTCGTTGACATCTGCTGCTGCTGCT 431
DB 361 ATTGGCTGTGACGCTACTCAACATACCGGTCCTATTCGTTGACATCTGCTGCTGCTGCT 420

QY 432 CGCGGATCTTCAGTGGCGCTGACCTCATCTCCACCTACTCTGTCGACCAAGAGCTGTGGT 491
DB 421 CGCGGATCTTCAGTGGCGCTGACCTCATCTCCACCTACTCTGTCGACCAAGAGCTGTGGT 480

QY 492 CGCGGCGCGCGCGCTCTTTTCGCGCGGAGCTTCATCGCCATCGTGTGCTGGCTACATCAGTA 551
DB 481 CGCGGCGCGCGCGCTCTTTTCGCGCGGAGCTTCATCGCCATCGTGTGCTGGCTACATCAGTA 551

```


BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AE003750: arm:3R [20671983,20899380]
estimated-cyto:96B10-96C1: 05/18/2001
Plate: GM.263 row: D column: 2
High quality sequence stop: 585.
Location/Qualifiers

FEATURES

1..694
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="GM26338"
/sex="female"
/dev_stage="newly eclosed females: germarium-stage 6"
/lab_host="X11 Blue"
/clone_lib="GM Drosophila melanogaster ovary pOT2"
/note="Organ: ovary; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2."
Location/Qualifiers

ORIGIN

Query Match 28.2%; Score 690.4; DB 4; Length 694;
Best Local Similarity 98.6%; Pred. No. 1e-173;
Matches 684; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1497 CCCAGCAGGATACGCGGTGAGTCCAACTGAGAGTATTGTTATTTGGCGTTCTAA 1556
DB 1 CCCAGCAGGATACGCGGTGAGTCCAACTGAGAGTATTGTTATTTGGCGTTCTAA 60

QY 1557 TGCTGTGATGATGTTGCTGTGTCACGACGTTGGTGACAGCAATGCTACTCCAGTC 1616
DB 61 TGCTGTGATGATGTTGCTGTGTCACGACGTTGGTGACAGCAATGCTACTCCAGTC 120

QY 1617 CCTCATTGCTTGGCTTCCAAACAGTCAAGATGGATCCCGCAACATTTAGACGATT 1676
DB 121 CCTCATTGCTTGGCTTCCAAACAGTCAAGATGGATCCCGCAACATTTAGACGATT 180

QY 1677 TCAGAGAGGCTTACTACTGCTTTCGAGAACACTGCGGATGATGCTCGGTTATGCTT 1736
DB 181 TCAGAGAGGCTTACTACTGCTTTCGAGAACACTGCGGATGATGCTCGGTTATGCTT 240

QY 1737 GGTGGGATTACGGATACAGATAGCGGGAATGCAACAGAGAGCTAGTGGATAATA 1796
DB 241 GGTGGGATTACGGATACAGATAGCGGGAATGCAACAGAGAGCTAGTGGATAATA 300

QY 1797 ATAGTGGAAACATAGTACATAGCGCTGTTGGCAAGCAATGCTTCAACCGAGGAGA 1856
DB 301 ATAGTGGAAACATAGTACATAGCGCTGTTGGCAAGCAATGCTTCAACCGAGGAGA 360

QY 1857 AGTCCTACGAAATATGACATCTCTTGACGTGACATACGTTTGGTGATCTTTGGCGGTG 1916
DB 361 AGTCCTACGAAATATGACATCTCTTGACGTGACATACGTTTGGTGATCTTTGGCGGTG 420

QY 1917 TGATCGGCTATTCTGGCGATGATATCAAGTTCTGTGGATGCTGCGAATGCTGAGG 1976
DB 421 TGATCGGCTATTCTGGCGATGATATCAAGTTCTGTGGATGCTGCGAATGCTGAGG 480

QY 1977 GAGAGCATCCCAAGGACATTAAGAAAGCGATTACTTTTACCAGCCCGGTGAATTCAGGG 2036
DB 481 GAGAGCATCCCAAGGACATTAAGAAAGCGATTACTTTTACCAGCCCGGTGAATTCAGGG 540

QY 2037 TAGATGCGGAAGTGCTCCGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACA 2096
DB 541 TAGATGCGGAAGTGCTCCGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACA 600

QY 2097 GATTTCGGGAATGAAGTTGGAATACAGAGGTCCTATCTGGATATGATCGCACACCTAACG 2156
DB 601 GATTTCGGGAATGAAGTTGGAATACAGAGGTCCTATCTGGATATGATCGCACACCTAACG 660

QY 2157 CCGTCATGGGAATTAAGGACTTCGATCTGACCTA 2190

DB 661 CCGTCATGGGAATTAAGGACTTCGATCTGACCTA 694

RESULT 12

CK657649

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

301	GGATTGTGGCGGTACCGTCTATCCCGGCTGTAGATTACGTCCGCGGAATCCATTGGC	360
378	TGCTGCACGTACTCAACATACCGGTCCATATTGGTGACATCTGCGTGTTCCTGGCGCCGA	437
361	TGCTGCACGTACTCAACATACCGGTCCATATTGGTGACATCTGCGTGTTCCTGGCGCCGA	420
438	TCCTTCAGTGGCTGACCTTCACATCTCCACCTACCTGCTGACCAAGGAGCTGTGCTCCGCGG	497
421	TCCTTCAGTGGCTGACCTTCACATCTCCACCTACCTGCTGACCAAGGAGCTGTGCTCCGCGG	480
499	GCGCGGGCTCTTTCGCGCGCAGCTTCAATCGCATCTGTGCGCTGGCTACATCATGAGTCGG	557
481	GCGCGGGCTCTTTCGCGCGCAGCTTCAATCGCATCTGTGCGCTGGCTACATCATGAGTCGG	540
558	TGGCTGGATCTGATCGATAACGAGGGCATTCGCCATATTTCGCCCTGCAAGTTCACCTACTTCC	617
541	TGCGTGGATCTGATCGATAACGAGGGCATTCGCCATATTTCGCCCTGCAAGTTCACCTACTTCC	600
618	TGTGGTGGCTCAGTGAGACTGGATCGGTGTTCTGTGCGCGCAGCGGCTTTGTGCT	677
601	TGTGGTGGCTCAGTGAGACTGGATCGGTGTTCTGTGCGCGCAGCGGCTTTGTGCT	660
678	ACTTCTACATGGTCTCCG	696
661	ACTTCTACATGGTCTCCG	679

RESULT 13	CK659064	linear	EST 30-JAN-2004
LOCUS	CK659064	682 bp	mRNA
DEFINITION	Lp16523.3 prime LP <i>Drosophila melanogaster</i> larval-early pupal POT2		
DEFINITION	<i>Drosophila melanogaster</i> cDNA clone Lp16523 5, mRNA sequence.		
ACCESSION	CK659064		
VERSION	CK659064.1	GI:41402589	
KEYWORDS	EST.		
SOURCE	<i>Drosophila melanogaster</i> (fruit fly)		
ORGANISM	<i>Drosophila melanogaster</i>		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; <i>Drosophila</i> .		
REFERENCE	1 (bases 1 to 682)		
AUTHORS	Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S. and Rubin, G.M.		
TITLE	BDGP/HMI <i>Drosophila</i> EST Project		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Stapleton, M.		
	BDGP		

FEATURES	ORIGIN
High quality sequence stop: 470.	
Location/Qualifiers	
1. .682	
/organism="Drosophila melanogaster"	
/mol_type="mRNA"	
/db_xref="taxon:7227"	
/clone="LP16523"	
/sex="male and female"	
/dev_stage="larvae-pupae"	
/lab_host="DH5-alpha"	
/clone_lib="LP Drosophila melanogaster larval-early pupal	
pot2"	
/note="organ: whole body; Vector: pot2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library."	
Query Match	28.0%; Score 676.2; DB 7; Length 682;
Best Local Similarity	99.6%; Pred. No. 1.4e-17;
Matches 678; Conservative	0; Mismatches 3; Indels. 0; Gaps 0;
LOCUS	
DEFINITION	
EST.	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
COMMENT	
BDGP	
Lawrence Berkeley National Lab	
BDGP/HEMI Drosophila EST Project	
Unpublished (2001)	
Contact: Stapleton, M.	
BDGP	
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.	
1 (bases 1 to 663)	
Ephydroidea; Drosophilidae; Drosophila.	
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
Drosophila melanogaster (fruit fly)	
Drosophila melanogaster	
EST.	
BG641172.1 GI:13773098	
BG641172	
04/13/2001, mRNA sequence.	
OstStt3: FBan0007748 ,enzyme, located on: 3R 96B16-96B17;	
pot2 Drosophila melanogaster cDNA clone S012448 5 similar to	
SD12448:5prime SD Drosophila melanogaster Schneider L2 cell culture	
linear	
EST 23-APR-2001	

One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit genomic AE003750: arm:3R [20671983,20899380]
 estimated-cyto:96B10-96C1; 04/13/2001 hit P element 1(3)J2D9:
 1(3)J2D9 A0026308 inserted at base 292 5' end of P element inverse
 PCR: 03/20/2001
 Plate: SD.124 row: D column: 12
 High quality sequence stop: 604.
 Location/Qualifiers

FEATURES

1..663
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="SD12448"
 /lab_hosts="DHS-alpha"
 /clone_lib="5D Drosophila melanogaster Schneider L2 cell
 culture pOT2"
 /note="Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized
 fractionated cDNAs were directly ligated into pOT2.
 Plasmid cDNA library."

ORIGIN

Query Match 27.3%; Score 659.8; DB 4; Length 663;
 Best Local Similarity 99.7%; Pred. No. 4.1e-168;
 Matches 661; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 TCTAAGCGAAGATGTCGTTCGATTCAGATTCGGTTATTAATTTTCGAGTTACTGGCTG 60
 1 TCTAAGCGAAGATGTCGTTCGATTCAGATTCGGTTATTAATTTTCGAGTTACTGGCTG 60
 61 GAATTGGGACATGAATCGGACCGCGAAGATCTGTAACAGCAAGTGGCTGGCTACAGCAG 120
 61 GAATTGGGACATGAATCGGACCGCGAAGATCTGTAACAGCAAGTGGCTGGCTACAGCAG 120
 121 CCTAATCACTTCGCCATCTCTGCTAATCGCTCGCTGGCGGATTTTCTCTCGCTCTT 180
 121 CCTAATCACTTCGCCATCTCTGCTAATCGCTCGCTGGCGGATTTTCTCTCGCTCTT 180
 181 CGCGTTCATCCGTTTCAGTCGATTAATTCATGATTTGATTCGGTTCATCTACCGGC 240
 181 CGCGTTCATCCGTTTCAGTCGATTAATTCATGATTTGATTCGGTTCATCTACCGGC 240
 241 CACCGCTTACATGCTGCGAGATGTTGGTACACTTCTCACTGGTTCGACGAGCGGC 300
 241 CACCGCTTACATGCTGCGAGATGTTGGTACACTTCTCACTGGTTCGACGAGCGGC 300
 301 ATGATATCCGCTCGGCGAGATGTTGGCGGTACCGTCTATCCGCGCTGATGATACGTC 360
 301 ATGATATCCGCTCGGCGAGATGTTGGCGGTACCGTCTATCCGCGCTGATGATACGTC 360
 361 CGGCGGAATCATTGGTGTGACGACGACGACGACGACGACGACGACGACGACGACGAC 420
 361 CGGCGGAATCATTGGTGTGACGACGACGACGACGACGACGACGACGACGACGACGAC 420
 421 CGTGTTCGCGCGGATCTTTCAGTGGCTGACCTCCATCTCCATCTCCATCTCCATCTCC 480
 481 CGAGCTGTGTCGCGCGGATCTTTCAGTGGCTGACCTCCATCTCCATCTCCATCTCCATCTCC 540
 481 CGAGCTGTGTCGCGCGGATCTTTCAGTGGCTGACCTCCATCTCCATCTCCATCTCCATCTCC 540
 541 CTATCATCAGTAGTGGTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 541 CTATCATCAGTAGTGGTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 601 GCAGTTCACTTCTCTGTTGGTGGCTCAGTGAAGACTGGATTCGTTCTGTTCTGGCTGGC 660
 601 GCAGTTCACTTCTCTGTTGGTGGCTCAGTGAAGACTGGATTCGTTCTGTTCTGGCTGGC 660
 661 CGC 663

Db 661 CGC 663

RESULT 15

BI357074

LOCUS

DEFINITION

666 bp mRNA linear EST 31-JUL-2001

REMARKS

Protophila melanogaster cDNA clone R43425 similar to Osetc3:

PBAN0007748 'enzyme' located on: 3R 96B16-96B17; 05/13/2001, mRNA

sequence.

ACCESSION

BI357074.1 GI:15051528

VERSION

EST.

KEYWORDS

Drosophila melanogaster (fruit fly)

SOURCE

ORGANISM

Drosophila melanogaster

Phylogeny: Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 666)

AUTHORS

Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,

Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,

George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,

Mista, S., Mungall, C.J., Nuncio, J., Pacleb, J., Paragas, V., Park, S.,

Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and

Rubin, G.M.

TITLE

BDGP/HMI RE Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit genomic AE003750: arm:3R [20671983,20899380]

estimated-cyto:96B10-96C1; 05/13/2001 hit P element 1(3)J2D9:

1(3)J2D9 A0026308 inserted at base 292 5' end of P element inverse

PCR: 05/13/2001

Plate: RE434 row: C column: 1

High quality sequence stop: 563.

Location/Qualifiers

1..666

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="RE43425"

/sex="male and female"

/dev_stages="0-24 hours mixed stage embryonic"

/lab_host="DHS-alpha T0A"

/clone_lib="RE Drosophila melanogaster normalized Embryo

pflc-1"

/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:

BamHI; Library was kindly generated by Piero Carninci at

the RIKEN. The library was normalized and excised using

Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 27.2%; Score 658; DB 4; Length 666;
 Best Local Similarity 100.0%; Pred. No. 1.3e-167;
 Matches 658; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TCTAAGCGAAGATGTCGTTCGATTCAGATTCGGTTATTAATTTTCGAGTTACTGGCTG 60
 9 TCTAAGCGAAGATGTCGTTCGATTCAGATTCGGTTATTAATTTTCGAGTTACTGGCTG 68
 61 GAATTGGGACATGAATCGGACCGCGAAGATCTGTAACAGCAAGTGGCTGGCTACAGCAG 120
 69 GAATTGGGACATGAATCGGACCGCGAAGATCTGTAACAGCAAGTGGCTGGCTACAGCAG 128
 121 CCTAATCACTTCGCCATCTCTGCTAATCGCTCGCTGGCGGATTTTCTCTCGCTCTT 180
 129 CCTAATCACTTCGCCATCTCTGCTAATCGCTCGCTGGCGGATTTTCTCTCGCTCTT 188
 181 CGCGGTCTATCCGTTTCAGTGGCTGATTCATGATTCGGTTCGTTCAACTACCGGC 240

```
Db ||||| 189 CCGGTCATCCGTTCCGAGTCGATTCATGAGTTGATCCGTTGTTCACTACCGGC 248
Qy ||||| 241 CACCGCTACATGGTGCAGATGGTGTGTAACCTCTCACTGGTTCGACGAGCGC 300
Db ||||| 249 CACCGCTACATGGTGCAGATGGTGTGTAACCTCTCACTGGTTCGACGAGCGC 308
Qy ||||| 301 ATGGTATCCGTCGCGGAGGATTCGCGGTACCGTCTATCCCGGCTGATTTAGTC 360
Db ||||| 309 ATGGTATCCGTCGCGGAGGATTCGCGGTACCGTCTATCCCGGCTGATTTAGTC 368
Qy ||||| 361 CCGCGGAATCCATTCGCTGCTGACGTAACATACCGGTCATATTCGTGACATCTG 420
Db ||||| 369 CCGCGGAATCCATTCGCTGCTGACGTAACATACCGGTCATATTCGTGACATCTG 428
Qy ||||| 421 CGTGTTCCTGGGCGCGATCTTCAGTGGCTGACCTCCATCCACCTACCTGACCAA 480
Db ||||| 429 CGTGTTCCTGGGCGCGATCTTCAGTGGCTGACCTCCATCCACCTACCTGACCAA 488
Qy ||||| 481 GGAGCTGTGCTCCGCGGCGCGGCTCTTCGCGGCGAGTTCAATCGGCTGG 540
Db ||||| 489 GGAGCTGTGCTCCGCGGCGCGGCTCTTCGCGGCGAGTTCAATCGGCTGG 548
Qy ||||| 541 CTACATCAGTAGTTCGCTGGTGGATCGTACGATAACGAGGGCATTCGCATATTCGCCCT 600
Db ||||| 549 CTACATCAGTAGTTCGCTGGTGGATCGTACGATAACGAGGGCATTCGCATATTCGCCCT 608
Qy ||||| 601 GCAGTTCACCTACTTCCTGTGGTGGCTGCTGAGTGAAGACTGGATCCGTTCTGGTCG 658
Db ||||| 609 GCAGTTCACCTACTTCCTGTGGTGGCTGCTGAGTGAAGACTGGATCCGTTCTGGTCG 666
```

Search completed: December 15, 2004, 11:09:49
Job time : 7356.2 secs

Blank
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 8574.94 Seconds
(without alignments)
4368.513 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 774

Sequence: 1 MRRTPMLNSKVAGYSSLLT.....RRKGYYIRNPVVVKGKRLK 774

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cygn2_1/USPTO_spool/US10028394/runat_14122004_131645_6800/app_query.fasta.1.3740
-DB=GenEmbl -QMT=fastap -SUFFIX=Oligo.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LISTN=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028394.SCGN.1.20262 @runat_14122004_131645_6800 -NCPU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_in:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	774	100.0	2417	3	AF132552 Drosophil
2	774	100.0	2417	6	AX799088 Sequence
3	774	100.0	2699	6	CQ589353 Sequence
4	464	59.9	4922	6	CQ589352 Sequence

C	5	464	59.9	162921	3	AC007853	Drosophil
C	6	464	59.9	181132	3	AC008206	Drosophil
C	7	464	59.9	227219	3	AE003750	Drosophil
C	8	390	50.4	75650	2	AC018145	Drosophil
C	9	208	26.9	2953	6	CQ596819	Sequence
C	10	83	10.7	2785	6	CQ596795	Sequence
C	11	47	6.1	54118	3	U13019	Caenorhabdi
C	12	42	5.4	433	6	AX340470	Sequence
C	13	42	5.4	507	6	AX886102	Sequence
C	14	42	5.4	507	6	BD025712	Sequence
C	15	42	5.4	764	6	AX136480	Sequence
C	16	42	5.4	764	6	BD123720	Sequence
C	17	42	5.4	787	6	AX869456	Sequence
C	18	42	5.4	787	6	BD149518	Primer fo
C	19	42	5.4	957	9	BC015880	Homo sapi
C	20	42	5.4	1097	5	AY437849	Ctenophar
C	21	42	5.4	1734	10	BC003206	Mus muscu
C	22	42	5.4	2284	6	AX882932	Sequence
C	23	42	5.4	2284	6	BD160013	Primer fo
C	24	42	5.4	2284	9	AX027789	Homo sapi
C	25	42	5.4	2481	9	AX799082	Sequence
C	26	42	5.4	2481	9	AY074880	Homo sapi
C	27	42	5.4	2503	6	AX017997	Sequence
C	28	42	5.4	2503	6	BD137303	Human nuc
C	29	42	5.4	2508	10	BC013054	Mus muscu
C	30	42	5.4	2510	6	AX136135	Sequence
C	31	42	5.4	2510	6	BD123520	Secretory
C	32	42	5.4	2510	9	AX075380	Homo sapi
C	33	42	5.4	2546	6	AX099510	Sequence
C	34	42	5.4	2546	6	BD063986	Secreted
C	35	42	5.4	2620	5	CR386955	Gallus ga
C	36	42	5.4	2710	6	AX799084	Sequence
C	37	42	5.4	4236	10	BC052433	Mus muscu
C	38	41	5.3	349	6	BD071886	Secreted
C	39	40	5.2	518	11	BY001913	Secreted
C	40	40	5.2	584	11	G89286	S208P6075
C	41	40	5.2	203492	2	AC107826	Mus muscu
C	42	40	5.2	227953	2	AC134292	Rattus no
C	43	40	5.2	285379	2	AC113070	Mus muscu
C	44	39	5.0	1040	6	CQ723424	Sequence
C	45	39	5.0	170691	9	AC104643	Homo sapi

ALIGNMENTS

RESULT 1	AF132552	Drosophila melanogaster	2417 bp	mRNA	linear	INV 12-FEB-2003
LOCUS	AF132552	Drosophila melanogaster	GM01838	full insert	CDNA.	
DEFINITION	AF132552	Drosophila melanogaster	GM01838	full insert	CDNA.	
ACCESSION	AF132552	Drosophila melanogaster	GM01838	full insert	CDNA.	
VERSION	AF132552.1	GI:4689327				
KEYWORDS	FLI-CDNA.					
SOURCE	Drosophila melanogaster (fruit fly)					
ORGANISM	Drosophila melanogaster					
REFERENCE	1 (bases 1 to 2417)					
AUTHORS	Tsang, G., Brokstein, P., Frise, E., Harvey, D., Evans-Holm, M., Lewis, S.E., Suh, C. and Rubin, G.M.					
TITLE	Direct Submission					
JOURNAL	Submitted (02-MAR-1999) Berkeley Drosophila Genome Project, University of California Berkeley, 539 Life Sciences Addition #3300, Berkeley, CA 94720, USA					
REFERENCE	2 (bases 1 to 2417)					
AUTHORS	Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Krommiller, B., Li, P., Liao, G., Miranda, A., Mungall, C.J., Nunco, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M. and Ceiniker, S.					
TITLE	Direct Submission					
JOURNAL	Submitted (12-FEB-2003) Berkeley Drosophila Genome Project, University of California Berkeley, 539 Life Sciences Addition #3300, Berkeley, CA 94720, USA					

Db	191	CGTTTCGAGTCGATTATCCATGAGTTTGATCGTGGTTCAACTACCGGGCCACCGCCTAC	250
Qy	61	MetValGlnAsnGlyTyrTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro	80
Db	251	ATGGTGCAGATGGTTGGTACAACTTCCTCAATGGTTTCGACGAGCGGCGATGGTATCCG	310
Qy	81	LeuGlyArgGluValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle	100
Db	311	CTGGCAGGATTGTGGGCGGTACCGTCTATCCGCGCTGATGATTACGTCCGCGGGAATC	370
Qy	101	HistPheLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120
Db	371	CAATTGGCTGTCGACGACTCAACATACCGGTCCATATTTCGTGACATCTGGGTGTCTCG	430
Qy	121	AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp	140
Db	431	GCSCCGATCTTCAGTGGCTGACCTCCATCTCCACTTCTGCTGCTGCTGCTGCTGCTG	490
Qy	141	SerAlaGlyValGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer	160
Db	491	TCCGCGGGCGCGCTCTTCGCGGCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT	550
Qy	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180
Db	551	AGTCTGGTGGCTGGATCGTACGATAACGAGGCGATTGCCATATTCCCTCTGAGTTACCC	610
Qy	181	TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAla	200
Db	611	TATTTCTCTGGTGGCTGGCTCAGTGAAGACTGGATCGTGTCTGGTGGCGCGCGCT	670
Qy	201	LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleLeuIle	220
Db	671	TTGTCTTACTCTACATCTCTGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG	730
Qy	221	ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer	240
Db	731	CCCTGTGACGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	790
Qy	241	TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe	260
Db	791	TACAGCACCTTCTACATCTCTGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG	850
Qy	261	GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla	280
Db	851	CAACCGATCGCACCGATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	910
Qy	281	ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe	300
Db	911	GTGGCCACCTTGGCCATTTGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	970
Qy	301	IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValValLeuThrMet	320
Db	971	ATCTGCGCGGATTTGCTGGGCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGGTGG	1030
Qy	321	LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla	340
Db	1031	CTGGCGCTTGTGGCGCGCTGGAGTGGAGCGCTTCTACTCGCTGGGAGTACTGGCTACGC	1090
Qy	341	LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer	360
Db	1091	AAGATCCACATTCCTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1150
Qy	361	PhePhePhePheLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle	380
Db	1151	TTCTTCTTTGATCTGACATCTCTGCTGGCTTCCAGTGGGAGTGTGCTGCTGCTGCTG	1210
Qy	381	LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe	400
Db	1211	AAGCAGATCAACGACGAGCGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1270
Qy	401	AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly	420
Db	1271	GCTGT	1330

COMMENT	Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA
COMMENT	Sequence submitted by:
COMMENT	Berkeley Drosophila Genome Project
COMMENT	Lawrence Berkeley National Laboratory
COMMENT	Berkeley, CA 94720
FEATURES	This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (http://fruitfly.berkeley.edu) or send email to cdna@fruitfly.berkeley.edu .
source	Location/Qualifiers
gene	1..2417
CDS	/organism="Drosophila melanogaster"
	/mol_type="mRNA"
	/db_xref="taxon:7227"
	<1..2417
	/genes="OstStt3"
	/db_xref="FLYBASE:FBgn0011336"
	<1..2395
	/genes="OstStt3"
	/note="Longest ORF"
	/codon_start=2
	/product="GMO1838p"
	/protein_id="A027851.2"
	/db_xref="GI:28317200"
	/db_xref="FLYBASE:FBgn0011336"
	/translation="LSEECVAFQIGNPRVTGMWDMNRTPKMLNSKVAGYSSLIITP AIIILWAGFSSRLFAVIRFESIHERPENRYRATAYMVONGWYNFLNFDERAWY PLGRIVGGTVFGLMTISGIIHLVLIPIVIRDCIYFLAPIESGLTSLTYLLTK ELMSAGAFIAFIAFVIGYISRSVAGSYDNEGAIAPALQTYLWRSYKTSVFN SAAALSYFMSAWGGYVFLINPLHVFLLIMGRYSPLLTSYFYLIGLLFSM QIPFGQPIRTSEHNAALGVFLLMVATLRLHLSVRSNEFRKLFIVGGLLVGVV FVAVVLTGVLWPNAGSRFYSLWDGTGAKIHIPITASVSEHQPTWFSFFDLHLV CAPGVVYCIKQINDERVFLVLAISAVVFAGVMRLMLTTPVVCMLAGVAFSGLL DVLOESSKNGTALSAATEVDEADSIEKTLVDKAGLKHRTKHDAODTGVSSN LKSIVILVLLMLMPFAVCTWTSNAYSPPSIVLAFHNSQDSNLIIDDFREAYVL SQNTADARVMSWDYQIAGVANRITLLVDNNTWNNSHIALYVKAMSTSEKSEYM TSLDVLVLFVFGVYGGDDINKFLWVRIAEGHPKDIKESDYFTDRGEFRVDAAE GAPALLNCLMYKLYRFGELKLDYRGPSPGYDRTNNAVIGNKDFDLTLLEBAYTTEHW LVRIYRKKPHEFNRPSTLKTERTIPPANFISRKNSKREKGYIRNRPVVVVGKRLK"
ORIGIN	
Alignment Scores:	
Pred. No.:	0
Score:	774.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	3
	Length: 2417
	Matches: 774
	Conservative: 0
	Mismatches: 0
	Indels: 0
	Gaps: 0
US-10-028-384-8 (1-774) x AFL132552 (1-2417)	
Qy	1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
Db	71 ATGAATCGGACGCGGAGAGTCTGAACAGCAAGTGGCTGGCTACAGACGCTTAATCAC 130
Qy	21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db	131 TTCGCCATCGTCGTAATCGCTGGCTGGCGGATTTTCTCTGCTGCTGCTGCTGCTGCTG 190
Qy	41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60

RESULT 2		AX799088	2417 bp	DNA	linear	PAT 08-OCT-2003
LOCUS		Sequence 7 from Patent WO03054008.				
DEFINITION		AX799088				
ACCESSION		AX799088				
VERSION		AX799088.1	GI:37605061			
KEYWORDS						
SOURCES		Drosophila melanogaster (fruit fly)				
ORGANISM		Drosophila melanogaster				
REFERENCE		1 Perreault, C. and McBride, K.				
AUTHORS		Mammalian SIMP protein, gene sequence and uses thereof in cancer				
TITLE		Patent: WO 03054008-A 7 03-JUL-2003;				
JOURNAL		Comptatigene Inc. (CA)				
FEATURES		Location/Qualifiers				
source		1..2417				
ORIGIN		/organism="Drosophila melanogaster"				
		/mol_type="genomic DNA"				
		/db_xref="taxon:7227"				
		/note="SFT3, Acc# AF132552"				
Alignment Scores:			Length:	2417		
Pred. No.:		0	Matches:	774		
Score:		774.00	Conservative:	0		
Percent Similarity:		100.00%	Mismatches:	0		
Best Local Similarity:		100.00%	Indels:	0		
Query Match:		100.00%	Gaps:	0		
DB:		6				
US-10-028-384-8 (1-774) x AX799088 (1-2417)						
QY	1	MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr	20			
Db	71	ATGATCGAGCGCGAGAGTGTGAACAGCAAGGTGGCTGGCTACAGCAGCTTAACACC	130			
QY	21	PhelalleLeuLeuLeuLeuAlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValle	40			
Db	131	TTCCGCATCTCTGCTAAATCGGCTGGCTGGCGGGATTTTCTCTCGCCTTTTCGGCGTCATC	190			
QY	41	ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr	60			
Db	191	CGTTTCGAGTCGATTATCCATGAGTTTGATCGGTGGTTCAACTACCGGCGCACCGCTTAC	250			
QY	61	MetValGlnAsnGlyTyrTyrAsnPhelLeuAsnTyrPheAspGluAlaTyrTyrPro	80			
Db	251	ATGTTGCAGAAATGTTGGTACAACTTCTCAACTGGTTTCGACGAGCGCGCATGTATCCG	310			
QY	81	LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle	100			
Db	311	CTCGCAGAGATGTTGGCGGTACCGTCTATCCCGGCTGATGATTAGTCCGCGGGAATC	370			
QY	101	HisTyrLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120			
Db	371	CATTGGCTGCTGCAGTACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCTTG	430			
QY	121	AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr	140			
Db	431	CGCGCATCTTCATGGCTGACCTCCATCTCCACTTACCTGCTGACCAAGAGGTGG	490			
QY	141	SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer	160			
Db	491	TCCGCGGCGCGCGCTCTTCGCGCGAGCTTCATCGCCATCGTGGCTGCTCATCAGT	550			
QY	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180			
Db	551	AGTTCGGTGGTGGATCGTACGATPACGAGGCAATTCGCATATTCGCCCTCGCATCACC	610			
QY	181	TyrPheLeuTyrValArgSerValIleThrGlySerValPheTyrSerAlaAlaAla	200			

Score:	774.00	Matches:	774
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-10-028-384-8 (1-774) x CQ589353 (1-2699)			
QY	1	MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr	20
DB	80	ATGAATCGACCCGCAAGATGCTGAACAGCAAGTGGCTGGCTACAGACGCTTAATCAAC	139
QY	21	PheAlaIleLeuLeuIleAlaTyrPheLeuAlaGlyPheSerSerArgLeuPheAlaValIle	40
DB	140	TTCCGCAATCTCTAATCGCGTGGCTGGCGGATTTTCCTCTCGCTCTTCGCGCTCATC	199
QY	41	ArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArgAlaThrAlaTyr	60
DB	200	CGTTTCGAGTGCATATCCATGAGTTTGCATCGGTTCACACTACCGCGGCACCGGCTAC	259
QY	61	MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTyrTyrPro	80
DB	260	ATGGTGCAGATGGTGGTACAACTTCCTCAACTGGTTCGACGAGCGCGATGGTATCCG	319
QY	81	LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle	100
DB	320	CTCGCAGGATTTGGCGGCTACCGTCTATCCCGCGCTGATGATTACGTCGCGGGAATC	379
QY	101	HisTrpLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120
DB	380	CATTGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATCTGGGTGTTCTG	439
QY	121	AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp	140
DB	440	GGCGCGAATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGACCAAGAGCTGTGG	499
QY	141	SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer	160
DB	500	TCCGCGGGCGCGGCTCTTCGCGCCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT	559
QY	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180
DB	560	AGGTGGTGGCTGGATCGTACATTAACGAGGCGCATGCCATATTTCGCCCTCGCAGTTCAC	619
QY	181	TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAlaAla	200
DB	620	TACTTCTGTGGTGGCTCAGTGAAGACTGGATCCGTTCTCTGGTGGCGCGCAGCGCT	679
QY	201	LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIle	220
DB	680	TTGTCTACTTCTACATGGTGTCCGCTGGGCTGGCTACGTTGTTTCATCAACACCTGATA	739
QY	221	ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer	240
DB	740	CCCCGTGCAGTCTTCGTACTGCTCATATAGGCGAGGTACTCGCGCGCTCTGCTGACCGC	799
QY	241	TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe	260
DB	800	TACAGCACCTTCTACATCTCGGACTGCTGTCTTCCATGCGAGATCCCTCTCGTGGATTC	859
QY	261	GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla	280
DB	860	CAACCGATACGACACAGTGAACACATGCTCGCTGGGTGGAGTGTGTGCTCTCTATGGCC	919
QY	281	ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe	300
DB	920	GTGGCCACCTTCGCGCATTTGAGTCCGCTGCTGGCGCAACGAGTTCGCGAAGCTGCTTC	979
QY	301	IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet	320
DB	980	ATCGTCGGCGGATTCGTGGTGGCGTTCGCTCTTTGTGGCGCTGCTGGTGGCTCACCATG	1039
QY	321	LeuGlyValAlaAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla	340

DB	1040	CTGGCGGTGTGGCGCGCTGGAGTGGAGCGCTTCTACTCGCTGGGATCTACTGCTACGCC	1099
QY	341	LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer	360
DB	1100	AAGATCCACATTCCTCATTCATTCGTCGTCGAGCATCAGCCACCATCTGCTTCTCG	1159
QY	361	PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle	380
DB	1160	TTCTTCTTTGATCTGCACATCTCTGGTGGCGCTTCCAGTGGGAGTGGTACTGATC	1219
QY	381	LysGlnIleAsnAspGluArgValPheValLeuTyrAlaIleSerAlaValTyrPhe	400
DB	1220	AAGCAGATCAACGACGAGCGCTTTCGTGGTGTGTACGCCATCAGTGGCGTTACTTC	1279
QY	401	AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly	420
DB	1280	CTGTGTGTGATGGTGGCTTTGATGTGACCTCACGCCGTGGTGTGCTGTCATGGCGGA	1339
QY	421	ValAlaPheSerGlyLeuLeuAspValPheLeuGlnLysAspSerSerLysArgMetGly	440
DB	1340	GTGGCTTTTCGGGACTGTTGGATGTCTCTCGACAGGATTCGCTTAAGCGAATGGGC	1399
QY	441	ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThr	460
DB	1400	ACAGCCATAAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTCATTTGAGAAAGACG	1459
QY	461	LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThr	480
DB	1460	CTGTACGACAAAGCTGGCAAGCTCAAGCATCGTACTAAGCATGATGCCACGAGGATCT	1519
QY	481	GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuMetMet	500
DB	1520	GGCGTCACTCCACCTCAAGAGATGATTGTTATTGTCCTTCTTAATGCTGTTGATGATG	1579
QY	501	PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu	520
DB	1580	TTGCTGTCTCACTGCACGTGGGTGACCCAGCAATGCTTACTCCAGTCCCTCCATGCTTG	1639
QY	521	AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyr	540
DB	1640	GCTTTCCACACAGTCAAGATGGATCCCGCAACAATTTAGACGATTTACAGAGGCTTAC	1699
QY	541	TyrTrpLeuSerGlnAsnThrAlaAspAlaArgValMetSerTrpTrpAspTyrGly	560
DB	1700	TACTGGCTTTCGACAGACACTGCCGATGATGCTCGCTTATGCTTGTGGGATTAACGA	1759
QY	561	TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn	580
DB	1760	TACCAGATAGCGGAATGGCAACAGCAACGCTAGTGGATTAATAATACGTGGAACAAT	1819
QY	581	SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle	600
DB	1820	AGTCACATAGCGCTGGTTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTACGAAAT	1879
QY	601	MetThrSerLeuAspValAspTyrValLeuValIlePheGlyValIleGlyTyrSer	620
DB	1880	ATGACATCTCTGACGTGGATACGTTTGGTGTATCTTTCGCGGTGTGATCGGCTATCT	1939
QY	621	GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys	640
DB	1940	GGCGATGATACCAACAGTTCCTGTGGATGGTCCGAATTCCTGAGGGAGAGCATCCCAAG	1999
QY	641	AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly	660
DB	2000	GACATTAAAGGAACGATTAATTTTACCGACCGCGGTAATTCAGGGGTAGATGCCGAAGT	2059
QY	661	AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu	680
DB	2060	GCTCGGCGCTGCTCAACTGCTTATGACAAATTAAGCTTACTACAGATTCGCGGATTC	2119
QY	681	LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn	700

Db 2120 AAGTTGGACTACAGAGTCCATCTGGATATGATCGCACACGTAACGCCGCTCATTCGGAA 2179

Qy 701 LysaspPheAspLeuThrTyrLeuGluGluAlaTyrThrGluHisTrpLeuValArg 720

Db 2180 AAGACTTCGATCTGACTACCTCGGAGAGGCTTACACACAGAACACCTGGCTTGTTCGC 2239

Qy 721 IleTyrArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740

Db 2240 ATCTATAGGTTGAAGAGCGCATGAGTTCAATAGACCATCACTGAAGACCAAGAGAGA 2299

Qy 741 ThrIleProProAlaAsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIle 760

Db 2300 ACGATTCTCCAGCAAACTTCATTTCGAGAAAGAACTCTAAGCGTCGCAAGGCTTACATA 2359

Qy 761 ArgAsnArgProValValValLysGlyLysArgThrLeuLys 774

Db 2360 CGAAACCGACCGGTTGTTGTTAAGGAAACGAACCTTGAAA 2401

RESULT 4

CQ589352/c

LOCUS CQ589352 4922 bp DNA linear PAT 02-FEB-2004

DEFINITION Sequence 17110 from Patent WO0171042.

ACCESSION CQ589352

VERSION CQ589352.1 GI:41648214

KEYWORDS Drosophila sp.

SOURCE Drosophila sp.

ORGANISM Drosophila sp.

REFERENCE 1 Venter, J.C., Adams, M., Li, P.W. and Myers, E.W. Detection kits, such as nucleic acid arrays, for detecting the expression of 10,000 or more Drosophila genes and uses thereof Patent: WO 0171042-A 17110 27-SEP-2001; PE Corporation (NY) (US)

TITLE Ephydroidea; Drosophilidae; Drosophila.

AUTHORS Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.

DATE 17110 27-SEP-2001

JOURNAL PE Corporation (NY) (US)

FEATURES

source 1..4922

location/Qualifiers

1..4922

/organism="Drosophila sp."

/mol_type="unassigned DNA"

/db_xref="taxon:7242"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4922

Score: 464.00 Matches: 464

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 59.95% Indels: 0

DB: 6 Gaps: 0

US-10-028-384-8 (1-774) x CQ589352 (1-4922)

Qy 1 MethAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20

Db 3843 ATGAATCCGACGCCGAAGATGCTCAACAGCAAGTGGCTGGCTACAGACGCTTAATCACC 3784

Qy 21 PheAlaIleLeuLeuIleAlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40

Db 3783 TTCGCCATCTGCTAATCGCTGGCTGGCGGATTTCTTCGCTCTTCGCCGCTCATC 3724

Qy 41 ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60

Db 3723 CGTTTCGAGTCGATATATCCATGATGTTGATCCGCTGGTTCAACTACCGGGCCACCGCCTAC 3664

Qy 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80

Db 3663 ATGGTGAGATAGTGTGGTACACTTCTCACTGGTTGAGAGCGGCGCATGGATCCG 3604

Qy 81 LeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100

Db 3603 CTCGGCAGGATTGTGGCGGTACCGTCTATCCCGGCTGATGATTACGTCCGGCGGAATC 3544

Qy 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120

Db 3543 CATGGCTGCTGCACGCTACTCAACATACGGTCCATATTCGTGACATCTCGGTGTTCTG 3484

Qy 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140

Db 3483 GGGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGAGTGTGG 3424

Qy 141 SerAlaGlyValAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160

Db 3423 TCCGGGGGGCGGGCTCTTCGCCCGCAGCTTCATCGCCATCGTGGTGGTACATCAGT 3364

Qy 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180

Db 3363 AGGTCCGTGGCTGGATCGTACGATTAACGAGGGCATTCGATATTCGCCCTGCGATTCACC 3304

Qy 181 TyrPheLeuTyrValArgSerValLysThrGlySerValPheThrSerAlaAlaAla 200

Db 3303 TACTTCTGTGGTGGCTCAGTGAAGACTGGATCCGCTGCTGGTGGCGCGACCGCT 3244

Qy 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIle 220

Db 3243 TTGTCTCTACTTCTACATGGTGGCTGGGTGGCTACGTGTTTCATCATCAACCTGATA 3184

Qy 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240

Db 3183 CCCTTCGACGCTTTCGTACTGCTCATTTATGGCAGGTACTCGCCGCTGCTGCTGACGACG 3124

Qy 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260

Db 3123 TACAGCACCTTCTACATCTGGACTGCTGTTCTTCATGCGAGATCCCTCTCGTGGGATTC 3064

Qy 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuMetAla 280

Db 3063 CAACCGATACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 3004

Qy 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300

Db 3003 GTGGCCACCTTCGCGCATTTTCGAGTCCGCTGCTGGCGCAACAGAGTTCGCGAAGCTGTT 2944

Qy 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320

Db 2943 ATCGTCGGCGGATTCGCTGGTGGCGCTGGCGCTCTTTGGCGCTGCTGGTGGCTGCTG 2884

Qy 321 LeuGlyValValAlaProTyrSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340

Db 2883 CTGGCGGTTGTGGCGGCTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 2824

Qy 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360

Db 2823 AGATCCACATTCCTCCATTCATTCGATCCGCTGGAGGATCAAGCCACCATCTGGTTCG 2764

Qy 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380

Db 2763 TTCTCTCTTGATCTGCACATCTCTGCTGGCTTCCAGTGGGAGTGGTGGTGGTGGTGG 2704

Qy 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400

Db 2703 AGCAGATCAACGACGAGCGGGTTCGTTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2644

Qy 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420

Db 2643 GCTGGT 2584

Qy 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440

Db 2583 GTGGCTTTCGGGACTGT 2524

Qy 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460

Db 2523 ACAGCCATTAAGCGGACCGACCGAAGTGGATGAAGTGGATTCATTTGAGAGAGACG 2464

Qy 461 LeuTyrAspLys 464

```

Db      2463 CIGTAGCAG 2452
|||||
RESULT 5
AC007853/c
LOCUS
DEFINITION
AC007853
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, J., An, H., Baldwin, D., Bauman, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Chao, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Fierstra, S., Frisoe, J., Gale, R.F., Garg, N.S., George, R.A.,
Gonzalez, C., Houch, J., Houskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalili, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, V., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Scheeler, F.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 3R, region 96B-96C
Unpublished
2 (bases 1 to 162921)
Celniker, S.E., Agbavani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Chao, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galic, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomtan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclet, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (17-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 27, 2001 this sequence version replaced gi:5670481.
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu.
FEATURES
Source
1..162921
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosomes="3R"
/map="96B-96C"
/clone="BACR03L02 (D766)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 162921
Score: 464.00 Matches: 464
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 59.95% Indels: 0
DB: 3 Gaps: 0
US-10-028-384-8 (1-774) x AC007853 (1-162921)
QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
Db 108481 ATCAATCGGACGCGGAGATGCTGACACAGCAGGTGGCTGGCTACAGCAGCTTAATCACC 108422
QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db 108421 TTCGCCATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGCCTCTTCGCGGTCATC 108362
QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
Db 108361 CGTTTCGATCGCATTTATCCATGAGTTTGATCCGTGGTTCAACTACCGGGCCACCCCTAC 108302
QY 61 MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
Db 108301 ATGGTCAGATGTTGGTACAACTTCCTCAACTGTTTCGACGAGCGCGCATGTTATCG 108242
QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db 108241 CTCGGCAGGATTTGTGGCGGATACCGCTATATCCCGCCTGATGATTACGTCGCGCGGAATC 108182
QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db 108181 CATTGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATCTGCGTGTCTCG 108122
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTrpLeuLeuThrIleGluLeuTrp 140
Db 108121 GCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCGTGG 108062
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db 108061 TCCGCGGCGCGCGGCTCTTTCGCGCGCAGCTTCATCGCATCTGCTGGCTGGCTACATCAGT 108002
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db 108001 AGGTCGGTGGTGGATCGTAGATAACGAGGCGATTGTCATATTCGCCCTCAGTTACC 107942
QY 181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSerAlaAlaAla 200
Db 107941 TACTTCCGTGGTGGCTGCTCAGTAGACTCGGCTGCTGCTGGCGCGCAGCGCT 107882
QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle 220
Db 107881 TTGTCCTACTTCTACATCGTGTCCGCTGGGCTGGCTAGCGTTTCATCATCACTGATA 107822
QY 221 ProLeuHisValPheValLeuLeuIleMetClyArgTyrSerProArgLeuLeuThrSer 240
Db 107821 CCCCTGACGCTTTCGTACTCTCATATGCGCAGGACTCGCGCGCTCTGCTGACGAGC 107762
QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
Db 107761 TACAGCACCTTCTACATCTCGGACTGCTGTCTCCATGAGATCCCTCTCGTGGGATTC 107702
QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
Db 107701 CAACCGATACGACACGACGACACATGCGCTGGGAGTGTGTGTGCTCTTATGGCC 107642
QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
Db 107641 GTGGCCACCTTTCGCCATTTGCGAGTCCGTGCTGTCGCGCAACGAGTTCCGGAAGCTGTC 107582
QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
Db 107581 ATCGTCGCGGATTCGTGGTGGGCGTTCGCGCTCTTTGTGGCCGCTCGTGGTGCACCATG 107522
QY 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340

```



```
Db 130700 CATTGGCTCTCCAGTACTCAACATACCGGTCCATATTCGTGACATCTGGTGTCTCTG 130641
Qy 121 AlaProillePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140
Db 130640 GGGCCGATCTTCAGTGGCTGAGCTCCATCTCCACCTACTCTGTCGACCAAGAGCTGTGG 130581
Qy 141 SerAlaGlyAlaGlyPheAlaAlaSerPheAlaAlaIleValProGlyTyrIleSer 160
Db 130580 TCCGGGGGGGGGGCTCTTCGCGCCAGCTTCATCGCCATCGTGGCTGTGCTACATCAGT 130521
Qy 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db 130520 AGTCCGTTGGTGGATCTGATGATACGAGGAGCTTCGCAATTCGCGCTCCAGTTCACC 130461
Qy 181 TyrPheLeuTyrValArgSerValGlyThrGlySerValPheThrSerAlaAlaAla 200
Db 130460 TACTTCCTGGTGGGCTGCTGCTGCGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 130401
Qy 201 LeuSerTyrPheTyrMetValSerAlaTyrGlyTyrValPheIleIleAsnLeuIle 220
Db 130400 TTGCTTACTTCTACATGCTGCTGCTGCGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCT 130341
Qy 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
Db 130340 CCCCCTGCAGCTCTTCGTACTCTCATTTATGGCAGGTACTCGCGCGCTGCTGCTGCTGCT 130281
Qy 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
Db 130280 TACAGCACCTTCACATCTGGGACTGCTGTCTCCATGAGATGCCCTTCGTGGGATTC 130221
Qy 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
Db 130220 CAACCGATACGACAGTGAACATGCTGCTGGCTGGAGTGTGTGCTCTTATGCTGCTGCT 130161
Qy 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
Db 130160 GTGGCCACCTTCGCGCATTCGAGTCCGCTGCTGCGGCAACGAGTTCCGGAAGCTGTTTC 130101
Qy 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
Db 130100 ATCGTCGCGGATGCTGGTGGGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 130041
Qy 321 LeuGlyValValAlaProThrSerGlyArgPheTyrSerLeuThrPheSerThrGlyTyrAla 340
Db 130040 CTGGGCGTTTGGCCCGCTGGAGTGGAGCTTCTACTCGCTGGGATACCTGGCTACGCC 129981
Qy 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrPheSer 360
Db 129980 AAGATCCACATTCACATCATGATCGTGTGGAGCATCAGCCACCATCTGGTTCCTCG 129921
Qy 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrTyrCysIle 380
Db 129920 TTCTTCTTGTATCTGCACATCTGCTGGTGGCTTCCAGTGGAGTGGTACTGCATC 129861
Qy 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
Db 129860 AACAGATCAACAGCAGCGCGTTTCGCTGGTGTCTGTACGCCATCAGTGGGTTTACTTC 129801
Qy 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
Db 129800 GCTGGTGTATGCTGGTGTGATGTTGACCCCTCACCGCGGTGGTGTGCTGCTGCTGCTGCT 129741
Qy 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerIleArgMetGly 440
Db 129740 GTGGCTTTTCGGGATGTTGGATGTTTCCTGCTGCAAGAGATTCGTTAAGGAAAGTGGC 129681
Qy 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysIleThr 460
Db 129680 ACAGCCATAAGCCACCCAGAGTGGATGATGATGATGATGATGATGATGATGATGATGATG 129621
Qy 461 LeuTyrAspLys 464
Db 129620 CTGTACGACAAG 129609
```

RESULT 8
AC018145/c
LOCUS AC018145 75650 bp DNA linear HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
ACCESSION AC018145
VERSION AC018145.1 GI:6553046
KEYWORDS HTG; HTGS PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 75650)
AUTHORS Adams, M. and Venter, J. C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:102.3542 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced.
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
1..75650
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 75650
Score: 390.00 Matches: 463
Percent Similarity: 99.78% Conservative: 0
Best Local Similarity: 99.78% Mismatches: 1
Query Match: 50.39% Indels: 1
DB: 2 Gaps: 0
US-10-028-384-8 (1-774) x AC018145 (1-75650)
Qy 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
Db 21487 ATGATCGAGCGCGGAGATGCTGAACAGCAAGTGGCTGGCTACAGCAGCTTAATCACC 21428
Qy 21 PheAlaIleLeuLeuIleAlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db 21427 TTCCGATCTCTGTAATCGCTGGCTGGCGGATTTCTCTCGCTCTTCGCGCTCATC 21368
Qy 41 ArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArgAlaThrAlaTyr 60
Db 21367 CGTTTCGATCGATATTCATGAGTTGATCCGTTGGTTTCAACTACCGGGCCACGCTAC 21308
Qy 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTyrTyrPro 80
Db 21307 ATGCTGCAAGATGTTGGTACAACTCTCTCAACTGGTTCGACGAGCGCATGTATCCG 21248
Qy 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db 21247 CTCGGCAGGATGTGGGCGGTACCGTCTATCCCGCGCTGTATGATTCGTCGGCGGAATC 21188
Qy 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db 21187 CATTTGGCTGTGACGCTACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCTG 21128
Qy 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140
Db 21127 GCGCGCATCTTCAGTGGCTTCGACCTCCATCTCCACCTACCTGTGTGACCAAGAGCTGTGG 21068
Qy 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db 21067 TCCGCGGGCGCGGCTCTTCGCGCGGCGAGCTTCATCGCCATCGTGGCTGCTACATCAGT 21008


```

DEFINITION      Sequence 24553 from Patent WO0171042.
ACCESSION       CQ596795
VERSION         CQ596795.1  GI:41653176
KEYWORDS
SOURCE          Drosophila sp.
ORGANISM        Drosophila sp.
                Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                Ephydroidea; Drosophilidae; Drosophila.
REFERENCE       1
AUTHORS         Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
TITLE           Detection kits, such as nucleic acid arrays, for detecting the
                expression of 10,000 or more Drosophila genes and uses thereof
                Patent: WO 0171042-A 24553 27-SEP-2001;
                PE Corporation (NY) (US)
FEATURES        location/Qualifiers
                1..2785
                /organism="Drosophila sp."
                /mol_type="unassigned DNA"
                /db_xref="taxon:7242"
ORIGIN
Alignment Scores:
Pred. No.:      1.21e-75      Length:      2785
Score:          83.00         Matches:      83
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    10.72%       Indels:        0
DB:              6           Gaps:          0

US-10-028-384-8 (1-774) x CQ596795 (1-2785)

Qy      603 SerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAsp 622
Db      2784 TCCTTGACGTGGACTACGTTTGGTGATCTTTGGCGGTGATCGGCTATTCTGGCAT 2725
Qy      623 AspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIle 642
Db      2724 GATATCAACAAGTTCCTGTGGATGTCGGAATGCTGAGGAGAGCATCCCAAGGACATT 2665
Qy      643 LysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaPro 662
Db      2664 AAGGAAGAGCATTAATTTACCGACCGCGGTGAATTCAGGATAGATCCGAAGTGCTCG 2605
Qy      663 AlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeu 682
Db      2604 GCCCTGCTCAACTGCCCTTATGTACAAATTAAGCTACTACAGATTCGGGATTTGAAGTTG 2545
Qy      683 AspTyrArg 685
Db      2544 GACTACAGG 2536

RESULT 11
UI3019
LOCUS          UI3019          54118 bp      DNA      linear      INV 13-JAN-2003
DEFINITION     Caenorhabditis elegans cosmid T12A2, complete sequence.
ACCESSION      UI3019
VERSION        UI3019.1  GI:912482
KEYWORDS       HTG.
SOURCE         Caenorhabditis elegans
ORGANISM       Caenorhabditis elegans
                Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
                Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
                1 (bases 1 to 54118)
                Waterston,R.
                Genome sequence of the nematode C. elegans: a platform for
                investigating biology. The C. elegans Sequencing Consortium
                Science 282 (5396), 2012-2018 (1998)
                99069613
                PUBLISHED 9851916
REFERENCE       2 (bases 1 to 54118)
AUTHORS        Latreille,P.
TITLE          The sequence of C. elegans cosmid T12A2

```

JOURNAL
REFERENCE
AUTHORS
TITLE

Unpublished (2001)
3 (bases 1 to 54118)
Waterston,R.
Direct Submission
Submitted (06-MAR-1996) Genome Sequencing Center, Washington
University

JOURNAL
REFERENCE
AUTHORS
TITLE

Unpublished (2001)
4 (bases 1 to 54118)
Waterston,R.
Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE

Submitted (26-JUN-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 54118)
Waterston,R.
Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE

Submitted (09-AUG-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
6 (bases 1 to 54118)
Waterston,R.
Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE

Submitted (19-APR-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
7 (bases 1 to 54118)
Waterston,R.
Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE

Submitted (28-AUG-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
8 (bases 1 to 54118)
Waterston,R.
Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE

Submitted (21-NOV-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
9 (bases 1 to 54118)
Waterston,R.
Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE

Submitted (13-JAN-2003) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
On Jul 27, 1995 this sequence version replaced gi:529354.
Submitted by:

COMMENT

Genome Sequencing Center
Department of Genetics, Washington University
St. Louis , MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one mi3 subclone.

For a graphical representation of this cosmid sequence and its
analysis see:

<http://www.wormbase.org/db/seq/sequence?name=T12A2;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C23G10, 200 bp overlap; the 3' cosmid is C18P10,
15000 bp overlap. Actual start of this cosmid is at base position 1
of T12A2; actual end is at 39122 of T12A2.

NOTES:

Coding sequences below are the result of integration and manual review of the following data : computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddb.jng.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans Orfome cloning project (<http://wofdb.dfci.harvard.edu/>), similarity to other proteins from BlastX analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. TRNAS are predicted using the program TRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

source

1. 54118
 /organism="Caenorhabditis elegans"
 /mol_type="genomic DNA"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="III"
 /clone="T12A2"
 3241. 6926
 /gene="hot-4"
 join(3241. 3365,3586. 3749,6392. 6561,6795. 6926)
 /gene="hot-4"
 /standard_name="T12A2.16b"
 /note="coded for by the following C. elegans cDNAs:
 yk1067e11.3, yk1325d06.3"
 /codon_start=1
 /product="Homolog of odr-2(two) protein 4, isoform b"
 /protein_id="AA15601.1"
 /db_xref="GI:20198879"
 /db_xref="WormBase:T12A2.16b"
 /translation="MRKKMTIVLCITLQATFSFLYHDSERKSLMRPKSHCFSCA
 SFVLPWSQLMHYPPNFTDRCPQSGIGLPCSSACFTLVERIDGVEYSEQH
 GVIRGMDRLLLGLDDVNNILSAYENQVCRHTDKLLRFLPLSGGTVVTFCSN
 GDFNCMDRLRELSNNSFOIHSVLLVATWILIF"
 join(3241. 3365,3586. 3740,6392. 6558,6795. 6926)
 /gene="hot-4"
 /standard_name="T12A2.16a"
 /note="coded for by the following C. elegans cDNAs:
 AF324056"
 /codon_start=1
 /product="Homolog of odr-2(two) protein 4, isoform a"
 /protein_id="AA15600.1"
 /db_xref="GI:20198878"
 /db_xref="WormBase:T12A2.16a"
 /translation="MRKKMTIVLCITLQATFSFLYHDSERKSLMRPKSHCFSCA
 SFVLPWSQLMHYPPNFTDRCPQSGIGLPCSSACFTLVERIDGVEYSEQH
 RGMCDRLLLGLDDVNNILSAYENQVCRHTDKLLRFLPLSGGTVVTFCSN
 NEHMDRLRELSNNSFOIHSVLLVATWILIF"
 complement(7287. 12051)
 /gene="T12A2.15"
 complement(join(7287. 7364,7412. 7520,7644. 7735,
 7844. 7979,8026. 8132,9511. 9663,9726. 9884,9932. 10111,
 10156. 10458,10598. 10687,10735. 10911,10953. 11156,
 11304. 11499,11832. 11919,11982. 12051))
 /gene="T12A2.15"
 /standard_name="T12A2.15a"
 /note="contains similarity to Pfam domain PF00168 (C2
 domain); coded for by the following C. elegans cDNAs:
 yk194a3.3, yk194a9.5, yk195a9.5, yk18e11.3, yk18e11.5"
 /codon_start=1
 /product="Hypothetical protein T12A2.15a"
 /protein_id="AAC24452.2"
 /db_xref="GI:20198877"
 /db_xref="WormBase:T12A2.15a"
 /translation="MWQSLVPLVGSAILSTFTFLGKWDYFVWVLIILIIASVTKS
 YLWKRERRILALRALREVEVIMAQLOLPFAWQVFPDTERVEVNLKVIQLWPYVG
 EYTKFNDFIIFQVKAQMPGNKFKFTKMDMGIPCRVGGIKYVTVNVRGRIIVD
 MDVAYAGDAFTVSCCGFTGMNNIQFSGKLRAILKPLPMPVGVGGVSTFLEMPKM
 DFNLTGGMENVLPGLDAIRSVINSQIAALCVLPNEIWPVLPAPDVDTQLYFFPEPDG

VVLKLIIEAKMLENEDISFIKKKSDPYAEIOVGSQSFKTETIDDDLNINWNEFEAV
 VQAQOQKLRILFPEDQKDEELGRSLVSDIKLVQAKGTIDKXVPLSGCKHGLHKA
 TWMNUSTEURLHLEKWEAEWGQADKPIHSAALLWYIDSVADLFPKSKLEPSPFVEV
 SLGKTQRTPVKTVNPLFQSKLFFVRLHGEQELFEAVDDGTRRSLSLIFLFT
 LIKENLEQONQMLTIGVQSPTVITRIMLRSDSQSLSHSGNSRSLRGLPFS
 KHEMKRETRADENGEIEIQIDFDLVNQLKIALICRDLMTDKKQDQNPYVSVKL
 VALDGNKEVFKKKTAKNTRKHFDNHEIDINPSDLLNHKVINVKDDTNYGTFA
 KVLGCLERLDSLNRLSQRWIPLSVERK"
 complement(join(7287. 7364,7412. 7520,7644. 7735,
 7844. 7979,8026. 8132,9386. 8532))
 /gene="T12A2.15"
 /standard_name="T12A2.15b"
 /note="contains similarity to Pfam domain PF00168 (C2
 domain); coded for by the following C. elegans cDNAs:
 yk349e10.5"
 /codon_start=1
 /product="Hypothetical protein T12A2.15b"
 /protein_id="AAC24451.1"
 /db_xref="GI:3258591"
 /db_xref="WormBase:T12A2.15b"
 /translation="MNELVSAFHIERANGKVGWVAVPAKTVNOAADDEVEIKLN
 VDNQMLRSDSQSLSHSGNSRLGRFGRSKHEMKRETRADENGEIEIQIDFDL
 VNLKIALICRDLMTDKKQDQNPYVSVKLVALDGNKEVFKKKTAKNTRPHFPN
 HVEIDINPSDLLNHKVINVKDDTNYGTFAKVLGCLERLDSLNRLSQRWIPLS
 VERK"

CDS

gene

CDS

Alignment Scores:

Pred. No.:	5,1e-37	Length:	54118
Score:	47.00	Matches:	47
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	6.07%	Indels:	0
DB:	3	Gaps:	0

US-10-028-384-8 (1-774) x UI3019 (1-54118)

Qy 546 AsnThrAlaAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGly 565

Db 33197 ACACAGCTGATGATGCCGCTTATGTCATGTTGGGATATGATATCAATCGCAGGA 33256

Qy 566 MetAlaAspArgThrThrLeuValAspAsnIleThrTrpAsnAsnSerHisIleAlaIleu 585

Db 33257 ATGGCAACCGTACTACATTAGTTGACAAACAATCCTGGACAACTCGCATTCGCCCTT 33316

Qy 586 ValGlyIleAlaMetSerSer 592

Db 33317 GTCCGAAAGCCATGCTCTCC 33337

RESULT 12

AX340470

LOCUS

DEFINITION

AX340470

ACCESSION

VERSION

AX340470.1

KEYWORDS

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

Jiang, Y., Harlocker, S.L. and Secrist, H.

Compcitions and methods for the therapy and diagnosis of colon

cancer

AX340470 433 bp DNA linear PAT 10-JAN-2002

Sequence 717 from Patent WO0196388.

AX340470

GI:18136452

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.

Jiang, Y., Harlocker, S.L. and Secrist, H.

Compcitions and methods for the therapy and diagnosis of colon

cancer

JOURNAL Patent: WO 0196388-A 717 20-DEC-2001;
 FEATURES CORIXA CORPORATION (US)
 source Location/Qualifiers
 1..433
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.52e-33 Length: 433
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-8 (1-774) x AX340470 (1-433)

QY 551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
 Db 26 GCACGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGAAT 85

QY 571 ThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 Db 86 ACGTTGGTGGATAATAACACCTGGAATAACAGCCACATAGCAGCTGGTGGGAAAAGCTATG 145

QY 591 SerSer 592
 Db 146 TCTTCT 151

RESULT 13
 AX886102
 LOCUS 507 bp DNA linear PAT 18-DEC-2003
 DEFINITION Sequence 1965 from Patent EP1033401.
 ACCESSION AX886102
 VERSION AX886102.1 GI:40042927
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 507)
 Edwards,J.B.D.M., Duclair,A. and Giordano,J.Y.
 Expressed sequence tags and encoded human proteins
 Patent: EP 1033401-A 1965 06-SEP-2000;
 Genset (FR)

FEATURES
 source Location/Qualifiers
 1..507
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 122..>505
 /notes="unnamed protein product"
 /codon_start=1
 /protein_id="CAF00240.1"
 /db_xref="GI:40042928"
 /translation="MSGWDYGVQIAGMANRTLVNNTWNNSHIALVGMSSNETAA
 YKIMRTLDYVLVIFGVIGVIGSGDDINKFLWVRVIAEGEHKPDIRESDYFPTQGEFR
 VDRAGSPTLLNCLMVKSVYRFGEMQ"

CDS
 Alignment Scores:
 Pred. No.: 1.73e-33 Length: 507
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-8 (1-774) x AX886102 (1-507)

QY 551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

ORIGIN
 Alignment Scores:
 Pred. No.: 1.73e-33 Length: 507
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-8 (1-774) x BD025712 (1-507)

QY 551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
 Db 113 GCACGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGAAT 172

QY 571 ThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 Db 173 ACGTTGGTGGATAATAACACCTGGAATAACAGCCACATAGCAGCTGGTGGGAAAAGCTATG 232

QY 591 SerSer 592
 Db 233 TCTTCT 238

RESULT 15
 AX136480
 LOCUS 764 bp DNA linear PAT 30-MAY-2001

Db 113 GCACGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGAAT 172
 QY 571 ThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 Db 173 ACGTTGGTGGATAATAACACCTGGAATAACAGCCACATAGCAGCTGGTGGGAAAAGCTATG 232
 QY 591 SerSer 592
 Db 233 TCTTCT 238

DEFINITION Sequence 402 from Patent EP1067182.
ACCESSION AX136480
VERSION AX136480.1 GI:14272884
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.
TITLE Secretory protein or membrane protein
JOURNAL Patent: EP 1067182-A 402 10-JAN-2001;
Helix Research Institute (JP)
FEATURES
source 1..764
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 2.44e-33 Length: 764
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 6 Gaps: 0
US-10-028-384-8 (1-774) x AX136480 (1-764)
Qy 551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaasnArgThr 570
Db 268 GCACGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGAACT 327
Qy 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
Db 328 ACCTTGGTGGATATACACCTGGATACAGCCACATAGCACTGGTGGGAAAGCTATG 387
Qy 591 SerSer 592
Db 388 TCTTCT 393
Search completed: December 14, 2004, 22:47:49
Job time : 8937.94 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 891.777 Seconds
(without alignments)
4556.130 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 774

Sequence: 1 MNRPEKMLNSKACVSSLI.....RRKGYIRNPVVKGRKTLK 774

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10028384/runat.14122004.131645.6790/app_query.fasta.1.3740
-DB=N Geneseq 23Sep04 -CPMT=fastap -SUFFIX=Oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=Digs -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10028384 @CEN 1.1.1986 @runat.14122004.131645.6790 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N Geneseq 23Sep04:
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	774	100.0	2417	10	ADD94789
2	774	100.0	2699	4	ABL13247
C 3	464	59.9	4922	4	ABL13246
C 4	208	26.9	2953	4	ABL18224
C 5	83	10.7	2785	4	ABL18208
6	42	5.4	433	6	ABL37128

ALIGNMENTS

RESULT 1
ADD94789
ID ADD94789 standard; DNA; 2417 BP.
XX
AC ADD94789;
XX
DT 29-JAN-2004 (first entry)
XX
DE Drosophila melanogaster STT3 gene sequence.
XX
KW source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
KW hematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide;
KW fruit fly; Gene; ds; STT3.
XX
OS Drosophila melanogaster.
XX
PN WO2003054008-A2.
XX
PD 03-JUL-2003.
XX
PF 18-DEC-2002; 2002WO-CA001967.
XX
PR 20-DEC-2001; 2001US-00028384.
XX
PA (COMP-) COMPATIGENE INC.

7	42	5.4	483	3	AAA43486	Aaa43486 Mouse sec
8	42	5.4	507	3	AAC01967	Aac01967 Human sec
9	42	5.4	754	5	AAF93968	Aaf93968 Primer bp
10	42	5.4	787	4	AAH07526	Aah07526 Human cDN
11	42	5.4	2284	4	AAH18021	Aah18021 Human cDN
12	42	5.4	2481	10	ADD94783	Add94783 Human STM
13	42	5.4	2510	5	AAF93772	Aaf93772 Human cDN
14	42	5.4	2537	4	AAD08289	Aad08289 Human sec
15	42	5.4	2546	2	AAV44866	Aav44866 Clone CTS
16	42	5.4	2546	5	AAF98843	Aaf98843 Human cDN
17	42	5.4	2547	4	AAD08315	Aad08315 Human sec
18	42	5.4	2660	6	ABQ54750	Abq54750 Human ova
19	42	5.4	2710	10	ADD94785	Add94785 Mouse SIM
20	41	5.3	349	2	AAV89737	Aav89737 EST clone
21	34	4.4	313	8	ABZ20328	Abz20328 Group III
22	34	4.4	387	4	ABA08337	Aba08337 Human sec
23	34	4.4	558	5	AAF93312	Aaf93312 Umbilical
24	34	4.4	1209	4	AAH33264	Aah33264 Human col
25	34	4.4	1209	6	ABL89850	Ab189850 Human pol
26	34	4.4	1543	2	AAX85055	Aax85055 Human sec
27	34	4.4	1543	8	ACD18981	Acd18981 Novel hum
28	34	4.4	1543	12	ADG78372	Adg78372 Human sec
29	34	4.4	1543	12	ADN60663	Adn60663 Human sec
30	34	4.4	1664	4	AAK94164	Aak94164 Human ful
31	34	4.4	1664	12	ADL30661	Adl30661 Full leng
32	32	4.1	1848	8	ABT18969	Abt18969 Aspergill
33	32	4.1	1969	8	ABT18375	Abt18375 Aspergill
34	32	4.1	2232	8	ABT20789	Abt20789 Aspergill
35	32	4.1	2603	8	ABT20191	Abt20191 Aspergill
36	32	4.1	3969	8	ABT17781	Abt17781 Aspergill
37	32	4.1	4603	8	ABT19595	Abt19595 Aspergill
38	30	3.9	1114	4	AAH99794	Aah99794 Human pro
39	28	3.6	281	3	AAA45221	Aaa45221 Human sec
40	28	3.6	500	6	ABV88332	Abv88332 Human col
41	28	3.6	616	3	AAZ80248	Aaz80248 Human col
42	27	3.5	474	5	AAF93633	Aaf93633 Umbilical
43	23	3.0	388	5	AAF66328	Aaf66328 Novel hum
44	22	2.8	2855	4	ABL02795	Ab102795 Drosophil
45	22	2.8	6153	4	ABL02794	Ab102794 Drosophil

XX PI Perreault C, McBride K;
 XX DR WPI; 2003-559122/52.
 XX DR P-PSDB; ADD94790.
 XX PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX PS Claim 6; SEQ ID NO 7; 66pp; English.
 XX CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the Drosophila melanogaster Srr3 gene which
 CC is related to the invention. Note: This sequence does not appear in the
 CC specification but was obtained by the indexer from GenBank.
 XX SQ Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 2417
 Score: 774.00 Matches: 774
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x ADD94789 (1-2417)

Qy 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20
 Db 71 ATGAATCGGACGGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGAGCTTAATCACC 130
 Qy 21 PheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValle 40
 Db 131 TTGCCCATCTGCTAATCGCTGGCTGGCGCGGATTTCTCTCGCTCTTCGCGTCATC 190
 Qy 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrAlaTyr 60
 Db 191 CGTTTCAGATCGATATATCCATAGATTTGATCCGTGGTTCAACTACCCGGGCCACCGCTAC 250
 Qy 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
 Db 251 ATGCTGCAGATGTTGGTACAACTTCTCAACTGGTTTCGACGAGCGCGATGGTATCGG 310
 Qy 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 311 CTCGGCAGGATTTGGGGCGGTACCGCTATCCCGGCGCTGATGATTACGTCGCGGGAATC 370
 Qy 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 371 CATTTGGCTGTCGACGTACTCAACATACCGGTTCATATTCGTGACATCTCGGTGTTCTCG 430
 Qy 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 Db 431 GCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 490
 Qy 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160

Db 491 TCCGGGGCGCGCGCTCTTTGCCGCCAGCTTTCGCCCATCGTGGCTGGCTAGTACGT 550
 Qy 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 Db 551 AGTTCGGTGGCTGGATCGTACGATAACGAGGCACTTGCATATTCGCCCTCGAGTTACCC 610
 Qy 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAla 200
 Db 611 TACTTCTCTGGTGGCTCGCTCAGTGAAGACTGGATCCGTGTTCTGTCGCGCGCGCCCT 670
 Qy 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle 220
 Db 671 TTGTCTTACTTCTACATGGTGTCCGCTGGGTGGCTAGTGTTCATCATCAACCTGATA 730
 Qy 221 ProLeuHisValPheValLeuLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSer 240
 Db 731 CCCTCGACGCTCTTCTGCTACTGCTTATTATGGGACGATCTCGCCGCTGTGTCGACGAGC 790
 Qy 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 Db 791 TAGACACCTTCTACATCTCGGACTGCTGTTCTCCATGCAGATCCCTTCGTGGGATTC 850
 Qy 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 Db 851 CAACCGATACGACACCATGAGTGGCTGGGTGGGAGTGTGTTGCTCTTATGGCC 910
 Qy 281 ValAlaThrLeuArgHisGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 Db 911 GTGGCACCTTGGCCCATTTGAGTCCGTGCTGTCGCGCAACGAGTTCCGAGAGCTGTC 970
 Qy 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuLeuThrMet 320
 Db 971 ATCGTGGCGGATTTGCTGGTGGCGTGGCTGCTTTGTGGCGCTGCTGGTGTCTCCCATG 1030
 Qy 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 Db 1031 CTGGCGTGTGGCCCCGTGGAGTGACCGCTTCTACTCGCTGGGAGTGTGTTACTCATC 1090
 Qy 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrTrpPheSer 360
 Db 1091 AAGATCCACATTCCTCATTCATCGGTGGGACATCAGCCACCATCTGGTTCGCTG 1150
 Qy 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
 Db 1151 TTCTCTTTGATCTGCACATCTCTGGTGGCGCTTCCAGTGGAGTGTGTTACTCATC 1210
 Qy 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 Db 1211 AAGCAGATCAACGACGAGCGGCTTTCGTGCTGTACGCCATCAGTGGCTTACTTC 1270
 Qy 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 Db 1271 GCTGGTGTGATGGTGGCTGTTGATGTTGACCTCACCGCTGGTGTGTCATCGTGGCGGA 1330
 Qy 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 Db 1331 GTGGCCCTTTTCGGGACTTGTGGATGTTCTCTGCACAGAGATTCGCTAAGCCGATGGGC 1390
 Qy 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
 Db 1391 ACAGCCATAGCGCAGCCACCGAAGTGGATGAAGTGGATTCATTCATTCAGAGAGACG 1450
 Qy 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThr 480
 Db 1451 CTGACGACAAAGGTGGCAAGCTGAAGCATCTGTAAGCATGATGCCGACGAGGATCT 1510
 Qy 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet 500
 Db 1511 GCGGTGACCTCCACCTGAAGAGATTTGTTATTGTCGCGCTTCTAATGCTGTTGATGATG 1570
 Qy 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu 520

Db 1571 TTCGCTGCCACTGCACGTGGGTGACAGCAATGCTACTCCACTCCCTCCATGCTTG 1630
 QY 521 AlapheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyr 540
 Db 1631 GCTTTCCACACAGTCAAGATCGATCCGCAACATTTTAGACGATTTTCAGAGAGGCTTAC 1690
 QY 541 TyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTpaAspTyrGly 560
 Db 1691 TACTGGCTTCCAGAACACATGCCGATGATGCTCGGTATGCTTGGTGGATACGGA 1750
 QY 561 TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 580
 Db 1751 TACCAGATAGCGGGAATGCAACAGACGCTAGTGGATAATAATACGTGGAACAA 1810
 QY 581 SerHisIleAlaLeuValGlyValAlaMetSerSerThrGluGluLysSerTyrGluIle 600
 Db 1811 ATCATATAGCTGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTACGAAT 1870
 QY 601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620
 Db 1871 ATGACATCTCTTGCCTGACTACCTGTTTGGTGTATCTTTGGCGGTGTGATCGGCTATCT 1930
 QY 621 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 640
 Db 1931 GCGCATGATATCAACAGTCTCTGTGGATGCTCGAATGCTGAGGAGAGCATCCCAAG 1990
 QY 641 AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAlaGluGly 660
 Db 1991 GACATTAAAGGAAGCGATTACTTTTACCAGCGCGGTGAATTCAGGGTAGATGCCAAGGT 2050
 QY 661 AlaProAlaLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680
 Db 2051 GCTCGCGCTCTCACTGCTGCTTATGACAAATTAAGCTACTACAGATTCCGGGAATTG 2110
 QY 681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
 Db 2111 AAGTTGGACTACAGAGGTCCATCTGCATATGATCGCACATGACGCGCTCAITGGGAAT 2170
 QY 701 LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArg 720
 Db 2171 AAGGACTTCGATCTGACTACTCTGGAGGAGGCTTACACACAGAACACTGCTTGTTCGC 2230
 QY 721 IleTyrArgValLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
 Db 2231 ATCTATAGGCTGAAGAAGCGCATGATTCATATAGACCATCATCTGAAGACCAAGGAGAG 2290
 QY 741 ThrIleProProAlaAsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIle 760
 Db 2291 AGATTCTCCAGCAAACTTCATTTCGAGAAGAACTCTAAGCGTCCGAAGGCTACATA 2350
 QY 761 ArgAsnArgProValValLysGlyLysArgThrLeuLys 774
 Db 2351 CGAAACCGACCGGTGTGTTAAGGGAACCAACCTTGAAC 2392

RESULT 2

ID ABL13247 standard; cDNA; 2699 BP.
 AC ABL13247;
 XX
 XX
 DT 26-MAR-2002 (first entry)
 XX
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 XX
 OS pharmaceutical; gene; ss.
 XX
 XX Drosophila melanogaster.
 XX
 PN W0200171042-A2.
 XX
 PD 27-SEP-2001.
 XX

PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PERE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR P-PSDB; ABB69144.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 34223; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB120511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins (AB557737-
 CC AB572072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2699
 Score: 774.00 Matches: 774
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x ABL13247 (1-2699)

QY 1 MethAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 Db 80 ATGAATCGACGCGCGAAGATGCTGACAGCAAGGTGGTGGTACAGCAGCTTAATCACC 139
 QY 21 PheAlaIleLeuLeuIleAlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 Db 140 TTCGCCATCTCTGTAATCGCTGGCTGGCGGATTTCTCTCGCTCTTCGCGCTAATC 199
 QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
 Db 200 CGTTTCGAGTCGATTATCCATGATTTGATCCGCTGGTTCACCTACCGGCGCACCGCTAC 259
 QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTyrPro 80
 Db 260 ATGGTGCAGAAATGGTGGTACAACTTCTCAACTGGTTCGACGAGCGCGATGGTATCCG 319
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 320 CTCGGCAGATTGTGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGCGGAATC 379
 QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 380 CATGGCTGCTGCAGCTACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCTG 439
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 Db 440 GCGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGAGCTGTG 499
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 Db 500 TCCGCGGCGCGCGCTCTTTCGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 559

PR 11-JUL-2000; 2000US-00614150.
 XX (PEXE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EM;
 XX WPI; 2001-65860/75.
 DR P-PSDB; ABB69143.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 34220; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB2072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 4922
 Score: 464.00 Matches: 464
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 59.95% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x ABL13246 (1-4922)

QY 1 MetAsnArgThrProIysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 DB 3843 ATGAATCGGACCGCCGAAGTGTGACACGACGAGTGGGTGGTCTACGACGCTAATCACC 3784
 QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 DB 3783 TTCGCCATCTCTAATCGCTGGCTGGCCGGAATTTCTCTCGCCTCTTCGCCGTATC 3724
 QY 41 ArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArgAlaThrAlaTyr 60
 DB 3723 CGTTTCGAGTCGATTATCCATGAGTTTGATCCGTGGTTCAACTACCGCGCCACCGCTAC 3664
 QY 61 MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluAlaTrpTyrPro 80
 DB 3663 ATGGTCACGATGGTTGGTACACTTCTCACTGGTTCGACGAGCGGCATGGTATCCG 3604
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 DB 3603 CTCGGCAGGATTTGGGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGGGAATC 3544
 QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 DB 3543 CATTGGCTGTCGACGATCTCAACATACCGTTCATATTCGTGACATCTGGGTCTCTG 3484
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 DB 3483 GCGCCGATCTCTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGTGG 3424
 QY 141 SerAlaGlyValGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 DB 3423 TCCGCGGGCGCCGCTCTTTCGCGCCGACCTTCATCGGCATCTGCTGGCTAGCATCAGT 3364
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 DB 3363 AGTTCGGTGGCTGGATCGTACGATACGAGGCGATTGGCATATTGCGCCTCAGTTACCC 3304

QY 181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSerAlaAlaAlaAla 200
 DB 3303 TACTTCTCTGTGGTGGCTCAGTGAAGACTGGATCCGTGTTCTGTCGCGCGAGCGCT 3244
 QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIle 220
 DB 3243 TTGTCCTACTTCTACATGGTGTCCGCTCGGGTGGCTACGTGTTCATCATCACTGATA 3184
 QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
 DB 3183 CCCCTGCACGCTTCTGACTGCTCATTTATGGCAGGTACTCGCGCGGTCTGCTGACGAC 3124
 QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 DB 3123 TACAGCACCCTCTACATCTCTGGACTGCTGTCTCCATGACAGATCCCTTCGTGGATTC 3064
 QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 DB 3063 CAACCGATACGACGACGATGACACATGCTGGCTGGAGTGTGTTGTGCTCTCTTATGGCC 3004
 QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 DB 3003 GTGGCCACCTTGGCCCATTTCCAGTCCGTGCTGTCGCGCACGAGTTCGGAAGCTGTT 2944
 QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 DB 2943 ATCGTGGCGGATGCTGCTGGTGGCGTGGGTCTCTTGTGGCGCTGCTGCTCACCATG 2884
 QY 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 DB 2883 CTGGGCGTGTGGCCCGCTGGAGTGGAGCGCTTCTACTCGCTGGATATCGCTAGCC 2824
 QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 DB 2823 AAGATCCACATTCGCATTCGCTGCGGAGCATCAGCCACCACTTGGTCTCG 2764
 QY 361 PhePhePheAsnLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
 DB 2763 TTCTTCTTTTATCTGCACATCTGTTGGTGGCTTCCAGTGGAGTGTGGTACTGCATC 2704
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 DB 2703 AAGCAGATCAACGACGAGCGGTTTCTGCTGTGTGTACGCCATCAGTGGCGTTACTTC 2644
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 DB 2643 GCTGT 2584
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 DB 2583 GTGGCCTTTTCGGGACTGTTGGATGTGTTCTGTCAGAGAGATTCGTCTAAGCGAATGGC 2524
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
 DB 2523 ACAGCCATAAGCGACGACCCAGCAAGTGGATGAAGTGGATTCATTTGAGAAAGAGACG 2464
 QY 461 LeuTyrAspLys 464
 DB 2463 CTGTACGACAG 2452

RESULT 4
 ABL18224/c
 ID ABL18224 standard; DNA; 2953 BP.
 XX
 AC ABL18224;
 DT 26-MAR-2002 (first entry)
 XX
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 6145.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 KW

```

XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PA 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 6145; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2953 BP; 924 A; 628 C; 648 G; 753 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9 41e-204 Length: 2953
Score: 208.00 Matches: 208
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.87% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) X ABL18224 (1-2953)
QY 1 MetAsnArgThrProLysMetLeuAsnSerIyeValAlaGlyTyrSerSerLeuLeuThr 20
DB 626 ATGAATCGAGCGGCGAAGATGCTGAACAGCAAGGTGGCTGGCTACAGCAGCCTAATCACC 567
QY 21 PheAlaIleLeuLeuIleAlaTTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
DB 566 TTGCCATCTGTAATCGCTGGCTGGCTGGCGGATTTCTCTCGCTCTTCGCGGTCATC 507
QY 41 ArgPheGluSerIleLeuHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
DB 506 CGATTTCAGTCGATTATCCATGATTTGATCCGCTGGTTCAACTACCGGCGCCACCGGCTAC 447
QY 61 MetValGlnAsnGlyTyrTrpAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
DB 446 ATGGTGCGAATGTTGGTGAACCTCTCACTGCTCACTGGTTCGACGAGCGCATGGTATCCG 387
QY 81 LeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
DB 386 CTCGGCAGGATTGTGGCGGTACCGCTATCCGGGCTGATGATTACGTCGCGCGGAATC 327
QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
DB 326 CATTTGGTGTGCGACGACTCAACATACCGGTCCATATTGTGACATCTCGTGTTCTCTG 267
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIysGluLeuTrp 140

```

```

DB 265 GCGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACTGCTGACCAAGAGCTGTGG 207
QY 141 SerAlaGlyAlaGlyLeuPheAlaIleAlaSerPheIleAlaIleValProGlyTyrIleSer 160
DB 206 TCCGGGGGGCGCGGCTCTTCGCCCGCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT 147
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
DB 146 AGTCTGGTGGCTGGATCGTACGATACGAGGGCATTCGCATATTCGCCCTCGAGTTACC 87
QY 181 TyrPheLeuTrpValArgSerValIysThrGlySerValPheTrpSerAlaAlaAla 200
DB 86 TACTTCTCTGGTGGCTGAGTCAAGACTGATCGTCTGTTCTGGTGGCGCGAGCCGCT 27
QY 201 LeuSerTyrPheTyrMetValSer 208
DB 26 TTGCTCTACTTCTACATGGTGICC 3

RESULT 5
ABL18208/c
ID ABL18208 standard; DNA; 2785 BP.
XX AC ABL18208;
XX XX
DT 26-MAR-2002 (first entry)
XX XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6097.
XX XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX XX pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PA 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 6097; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2785 BP; 795 A; 644 C; 604 G; 742 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8 3e-75 Length: 2785
Score: 83.00 Matches: 83
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.72% Indels: 0

```

DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x ABL37128 (1-2785)

QY 603 SerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAsp 622
 DB 2784 TCTCTGACGTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTTCTGGGAT 2725

QY 623 AspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIle 642
 DB 2724 GATATCAACAAGTTCCTGTGGATGTCGGAATTCGTAGGAGAGCATCCCAAGACATT 2665

QY 643 LysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyValPro 662
 DB 2664 AAGGAACCGATTACTTTTACCGACCGCGGTGAATTCAGGGTAGATGCCAAGGTGCTCG 2605

QY 663 AlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeu 682
 DB 2604 GCCCTGCTCACTGCTTATGTACAAATTAGCTACTACATTCCGGGAATTGAAGTTG 2545

QY 683 AspTyrArg 685
 DB 2544 GACTACAGG 2536

RESULT 7

ABL37128
 ID ABL37128 standard; cDNA; 433 BP.
 AC ABL37128;
 XX
 DT 08-APR-2002 (first entry)
 DE Human colon tumour antigen polynucleotide SEQ ID NO:717.
 DE Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
 KW Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
 KW colon tumour metastatic antigen; diagnosis; gene; ss.
 XX Homo sapiens.
 OS
 XX
 XX WO2001:96388-A2.
 XX
 XX 20-DEC-2001.
 XX
 XX 08-JUN-2001; 2001WO-US018557.
 XX
 XX 09-JUN-2000; 2000US-0210899P.
 XX
 XX 20-FEB-2001; 2001US-0270216P.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Jiang Y, Harlocker SL, Secrist H;
 XX WPI; 2002-114514/15.
 XX
 XX Novel isolated colon tumor polynucleotide differentially expressed in
 PT colon tumor or colon metastatic tumor and polypeptides encoded by them,
 PT useful for inhibiting development of cancer in patient.
 XX
 XX Claim 1; SEQ ID NO 717; 105pp; English.
 PS
 XX ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (1)
 CC which were isolated from human colon tumour and colon metastatic tumour
 CC cDNA libraries. (1) have cytostatic activity and can be used in vaccine
 CC production. (1) can be used for stimulating and/or expanding T cells
 CC specific for a tumour protein on contact with the T cells. They are also
 CC useful for inhibiting the development of cancer in a patient. (1) can be
 CC used as probes or primers for nucleic acid hybridisation, for preparing
 CC mutant species primers, or primers for use in genetic constructions. (1)
 CC can be used in the diagnosis of a colon tumour

US-10-028-384-8 (1-774) x ABL37128 (1-433)

QY 551 AlaArgValMetSerTyrTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
 DB 26 GCACGAGTAAATGCTTGGTGGGATTATGCTATCAGATAGCTGGAATGGCTAATAGAACT 85

QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 DB 86 ACCTTGGTGGATATATACACCTGGGAATACACCCATAGCACTGGTGGGAAAGCTATG 145

QY 591 SerSer 592
 DB 146 TCTTCT 151

RESULT 7

AAA43486
 ID AAA43486 standard; cDNA; 483 BP.
 AC AAA43486;
 XX
 DT 21-AUG-2000 (first entry)
 DE Mouse secreted expressed sequence tag SEQ ID NO:61.
 DE Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antistimulant; vulnerable; antiparkinsonian;
 KW antitumor; osteopathic; neuroprotective; nontoxic; antiparasitic;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;
 KW autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
 KW infection; depression; psoriasis; ss.
 XX
 XX Mus musculus.
 OS
 XX WO200021991-A1.
 XX
 XX 20-APR-2000.
 XX
 XX 15-OCT-1999; 99WO-US024206.
 XX
 XX 15-OCT-1998; 98US-0104436P.
 XX
 XX (GENY) GENETICS INST INC.
 XX
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR;
 PI WPI; 2000-317938/27.
 DR
 XX Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (sESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders.
 XX
 XX Claim 1; Page 217; 803pp; English.
 PS
 XX AAA43426 to AAA45925 represent specifically claimed secreted expressed
 CC sequence tags (sESTs), isolated from human, mouse, chicken and rat tissue
 CC sources. The sESTs can have a range of activities depending on the
 CC tissues they were isolated from. The activities include: chemotactic;

Alignment Scores:

CC proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
CC haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
CC antifungal; antiviral; antidiabetic; antiaesthetic; vulnery; antitumor;
CC osteopathic; neuroprotective; nootropic; antiparkinsonian; antipsoriatic;
CC cerebroprotective; anticonvulsant; and antidiabetic. The ESTs can be
CC used for gene therapy and in vaccines. The ESTs are useful as probes for
CC the identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the ESTs. Proteins encoded by the ESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma, myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention

XX SQ Sequence 483 BP; 123 A; 115 C; 128 G; 117 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,14e-33 Length: 483
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-8 (1-774) x AAA43486 (1-483)

QY 551 AlaArgValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
Db 196 GCCCGGTCATGTCGTTGGGACACGCTATCAGATGCTGGCATGGCCACAGGACC 255
QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
Db 256 ACTCTGGTGGATTAACACACCTGGAAACACAGCCACATCGCACTGGTGGGAAAGCTATG 315
QY 591 SerSer 592
Db 316 TCTTCT 321

RESULT 8
AAC01967
ID AAC01967 standard; cDNA; 507 BP.

XX AAC01967;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 1965.

XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX Gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

XX (GIST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG01961.

XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 1; SEQ ID NO 1965; 71pp + Sequence Listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors

XX SQ Sequence 507 BP; 158 A; 85 C; 118 G; 146 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,28e-33 Length: 507
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-8 (1-774) x AAC01967 (1-507)

QY 551 AlaArgValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
Db 113 GCACGAGTAATGCTTGTGGGATTATGCTATCAGATAGCTGAATGCTAATAGAACT 172
QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
Db 173 ACCTTGGTGGATAATAACACCTTGAATAACAGCCATAGCACTGGTGGGAAAGCTATG 232
QY 591 SerSer 592
Db 233 TCTTCT 238

RESULT 9
AAF93968
ID AAF93968 standard; DNA; 764 BP.

XX AAF93968;

XX 23-MAY-2001 (first entry)

XX Primer specific for DNA encoding secretory/membrane protein SEQ ID 402.

XX Human; secretory protein; membrane protein; vaccine; gene therapy;
XX rheumatoid arthritis; diabetes; PCR primer; ss.

XX Synthetic.

XX EP1067182-A2.

XX 10-JAN-2001.

XX 07-JUL-2000; 2000EP-00114090.

XX 08-JUL-1999; 99JP-00194179.

XX 11-JAN-2000; 2000JP-00118775.

XX 02-MAY-2000; 2000JP-00183766.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

XX DR WPI; 2001-093989/11.

XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in

XX PT gene therapy or as candidate target molecules in drug development.

XX PS Claim 4; SEQ ID NO 402; 609pp + Sequence Listing; English.

XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916

XX CC which encode human secretory or membrane proteins represented by AAB88317

XX CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and

XX CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the

XX CC invention. The invention also includes methods for the production of

XX CC antibodies directed against the proteins, and cDNA sequences, which can

XX CC be used in vaccines. The polynucleotide sequences can be used in gene

XX CC therapy. The polynucleotide sequences and the proteins they encode may be

XX CC used in the prevention, treatment and diagnosis of diseases associated

XX CC with inappropriate secretory protein/membrane protein expression. The

XX CC nucleic acids and complementary sequences may also be used as DNA probes

XX CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect

XX CC and quantitate the presence of similar nucleic acid sequences in samples.

XX CC They may also be used to study the expression and function of secretory

XX CC proteins/membrane polypeptides and their role in metabolism. The

XX CC polypeptides may be used as antigens in the production of antibodies

XX CC against them and in assays to identify modulators (agonists and

XX CC antagonists) of expression and activity. The antibodies and antagonists

XX CC may also be used as therapeutic agents to down regulate expression and

XX CC activity. The antibodies may also be used as diagnostic agents for

XX CC detecting the presence of the polypeptides in samples (e.g. by enzyme

XX CC linked immunosorbent assay (ELISA). Examples of diseases which may be

XX CC treated include rheumatoid arthritis and diabetes

XX SQ Sequence 764 BP; 244 A; 131 C; 169 G; 214 T; 0 U; 6 Other;

Alignment Scores:

Pred. No.:	4,85e-33	Length:	764
Score:	42.00	Matches:	42
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.43%	Indels:	0
DB:	5	Gaps:	0

US-10-028-384-8 (1-774) x AAF93968 (1-764)

QY 551 AlaArgValMetSerTrpTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

DB 268 GCACGAGTAATGCTCTGGTGGGATTATGGCTATCAGATAGCTGGATGGCTATAGAACT 327

QY 571 ThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590

DB 328 ACGTTGGTGGATAATAACACCTGGGAATACAGCCACATAGCACTGGTGGGAAGACTATG 387

QY 591 SerSer 592

DB 388 TCTTCT 393

RESULT 10

AAH07526

ID AAH07526 standard; cDNA; 787 BP.

XX AC AAH07526;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA clone (5'-primer) SEQ ID NO:4361.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX XX 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-00116126.

XX PR 29-JUL-1999; 99JP-00248036.

XX PR 27-AUG-1999; 99JP-00300253.

XX PR 11-JAN-2000; 2000JP-00118776.

XX PR 02-MAY-2000; 2000JP-00183767.

XX PR 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX PT length cDNAs defined in the specification, and for the detection and/or

XX PT diagnosis of the abnormality of the proteins encoded by the full-length

XX PT cDNAs.

XX PS Claim 1; SEQ ID NO 4361; 2537pp + Sequence Listing; English.

XX CC The present invention describes primer sets for synthesizing 5602 full-

XX CC length cDNAs defined in the specification. Where a primer set comprises:

XX CC (a) an oligo-dT primer and an oligonucleotide complementary to the

XX CC complementary strand of a polynucleotide which comprises one of the 5602

XX CC nucleotide sequences defined in the specification, where the

XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX CC of an oligonucleotide comprising a sequence complementary to the

XX CC complementary strand of a polynucleotide which comprises a 5'-end

XX CC sequence and an oligonucleotide comprising a sequence complementary to a

XX CC polynucleotide which comprises a 3'-end sequence, where the

XX CC oligonucleotide comprises at least 15 nucleotides and the combination of

XX CC the 5'-end sequence/3'-end sequence is selected from those defined in the

XX CC specification. The primer sets can be used in antisense therapy and in

XX CC gene therapy. The primers are useful for synthesizing polynucleotides,

XX CC particularly full-length cDNAs. The primers are also useful for the

XX CC detection and/or diagnosis of the abnormality of the proteins encoded by

XX CC the full-length cDNAs. The primers allow obtaining of the full-length

XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893

XX CC represent human amino acid sequences; and AAH13629 to AAH13632 represent

XX CC oligonucleotides, all of which are used in the exemplification of the

XX CC present invention

XX SQ Sequence 787 BP; 252 A; 137 C; 168 G; 220 T; 0 U; 10 Other;

Alignment Scores:

Pred. No.:	4.99e-33	Length:	787
Score:	42.00	Matches:	42
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.43%	Indels:	0
DB:	4	Gaps:	0

US-10-028-384-8 (1-774) x AAH07526 (1-787)

QY 551 AlaArgValMetSerTrpTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

DB 177 GCACGAGTAATGCTCTGGTGGGATTATGGCTATCAGATAGCTGGATGGCTATAGAACT 236

QY 571 ThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590

DB 237 ACGTTGGTGGATAATAACACCTGGGAATACAGCCACATAGCACTGGTGGGAAGACTATG 296

QY 591 SerSer 592

DB 297 TCTTCT 302

RESULT 11

AAH18021

ID AAH18021 standard; cDNA; 2284 BP.

XX AC AAH18021;
 XX DT 26-JUN-2001 (first entry)
 XX DE Human cDNA sequence SEQ ID NO:17837.
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX OS Homo sapiens.
 XX PN EF1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-00116126.
 XX PR 29-JUL-1999; 99JP-00248036.
 XX PR 27-AUG-1999; 99JP-00300253.
 XX PR 11-JAN-2000; 2000JP-00118776.
 XX PR 02-MAY-2000; 2000JP-00183767.
 XX PR 09-JUN-2000; 2000JP-00241899.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX DR WPI; 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 XX PS Claim 8; SEQ ID NO 17837; 2537pp + Sequence Listing; English.
 XX CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92445 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 XX SQ Sequence 2284 BP; 726 A; 388 C; 436 G; 734 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,39e-32 Length: 2284
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-8 (1-774) x AAH18021 (1-2284)
 QY 551 AlaArgValMetSerTTPAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

Db 177 GCACGAGTAATGCTCTGGTGGGATATGCTATCATAGCTAGCTAGGCTAATAGAACT 236
 QY 571 ThrLeuValAspAsnAsnThrTTPAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 Db 237 ACCTTGGTGGATAAATAACACCTTGGATAACACCAATAGCACTGCTGGGAAAGCTATG 296
 QY 591 SerSer 592
 Db 297 TCCTCT 302
 RESULT 12
 ADD94783
 ID ADD94783 standard; cDNA; 2481 BP.
 XX AC ADD94783;
 XX DT 29-JAN-2004 (first entry)
 XX DE Human SIMP cDNA sequence.
 XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
 KW gene; ss.
 XX OS Homo sapiens.
 XX PN W02003054008-A2.
 XX PD 03-JUL-2003.
 XX PF 18-DEC-2002; 2002WO-CA001967.
 XX PR 20-DEC-2001; 2001US-00028384.
 XX PA (COMP-) COMPATIGENE INC.
 XX PI Perreault C, McBride K;
 XX DR WPI; 2003-559122/52.
 XX DR P-PSDB; ADD94784.
 XX PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.
 XX PS Claim 6; SEQ ID NO 1; 66pp; English.
 XX CC This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC)-associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestine cancer, sarcoma, prostate cancer, testicular cancer, breast cancer, melanoma, pancreatic cancer or haematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the human SIMP cDNA which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from GenBank.

SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.49e-32 Length: 2481
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x ADD94783 (1-2481)

QY 551 AlaArgValMetSerTrpTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
 Db 1795 GCACGAGTAATGCTTGGTGGATATGGCTATCAGATAGCTGAATGGCTAATAGAACT 1854

QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 Db 1855 ACGTTGGTGGATAAATACACCTGGATAACAGCCACATAGCACTGGTGGAAAGCTATG 1914

QY 591 SerSer 592
 Db 1915 TCTTCT 1920

US-10-028-384-8 (1-774) x AAF93772 (1-2510)

QY 551 AlaArgValMetSerTrpTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
 Db 268 GCACGAGTAATGCTTGGTGGATATGGCTATCAGATAGCTGAATGGCTAATAGAACT 327

QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 Db 328 ACGTTGGTGGATAAATACACCTGGATAACAGCCACATAGCACTGGTGGAAAGCTATG 387

QY 591 SerSer 592
 Db 388 TCTTCT 393

RESULT 14
 AAD08289
 ID AAD08289 standard; cDNA; 2537 BP.
 XX AC AAD08289;
 XX DT 08-AUG-2001 (first entry)
 XX DE Human secreted protein-encoding gene 7 cDNA clone HDTLR06, SEQ ID NO: 17.
 XX KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;
 KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
 KW cell culture; chemotaxis; vulnerability; binding partner identification;
 KW gene therapy; ss.
 XX OS Homo sapiens.
 XX PH Key Location/Qualifiers
 FT CDS 93..953
 FT /tag= a
 FT /product= "Human secreted protein precursor"
 FT sig_peptide 93..155
 FT /tag= b
 FT mat_peptide 156..950
 FT /tag= c
 FT /product= "Mature human secreted protein"
 XX WO200136440-A1.
 XX 25-MAY-2001.
 XX

SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.49e-32 Length: 2481
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x ADD94783 (1-2481)

QY 551 AlaArgValMetSerTrpTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
 Db 1795 GCACGAGTAATGCTTGGTGGATATGGCTATCAGATAGCTGAATGGCTAATAGAACT 1854

QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 Db 1855 ACGTTGGTGGATAAATACACCTGGATAACAGCCACATAGCACTGGTGGAAAGCTATG 1914

QY 591 SerSer 592
 Db 1915 TCTTCT 1920

US-10-028-384-8 (1-774) x AAF93772 (1-2510)

QY 551 AlaArgValMetSerTrpTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
 Db 268 GCACGAGTAATGCTTGGTGGATATGGCTATCAGATAGCTGAATGGCTAATAGAACT 327

QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 Db 328 ACGTTGGTGGATAAATACACCTGGATAACAGCCACATAGCACTGGTGGAAAGCTATG 387

QY 591 SerSer 592
 Db 388 TCTTCT 393

RESULT 13
 AAF93772
 ID AAF93772 standard; cDNA; 2510 BP.
 XX AC AAF93772;
 XX DT 23-MAY-2001 (first entry)
 XX DE Human cDNA encoding a membrane or secretory protein clone PSECO070.
 XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes; ss.
 XX OS Homo sapiens.
 XX PF EF1067182-A2.
 XX PD 10-JAN-2001.
 XX PF 07-JUL-2000; 2000EP-00114090.
 XX PR 08-JUL-1999; 99JP-00194179.
 XX PR 11-JAN-2000; 2000JP-00118775.
 XX PR 02-MAY-2000; 2000JP-00183766.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 XX WPI; 2001-093989/11.
 XX DR P-PSDB; AAB88345.
 XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 XX gene therapy or as candidate target molecules in drug development.
 XX PS Claim 1; SEQ ID NO 57; 609pp + Sequence Listing; English.
 XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by AAB88317
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
 CC AAF6232 - AAF62235 which are used to isolate the cDNA sequences of the
 CC invention. The invention also includes methods for the production of
 CC antibodies directed against the proteins, and cDNA sequences, which can
 CC be used in vaccines. The polynucleotide sequences can be used in gene
 CC therapy. The polynucleotide sequences and the proteins they encode may be
 CC used in the prevention, treatment and diagnosis of diseases associated
 CC with inappropriate secretory protein/membrane protein expression. The
 CC nucleic acids and complementary sequences may also be used as DNA probes
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect

CC and quantitate the presence of similar nucleic acid sequences in samples.
 CC They may also be used to study the expression and function of secretory
 CC proteins/membrane polypeptides and their role in metabolism. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC against them and in assays to identify modulators (agonists and
 CC antagonists) of expression and activity. The antibodies and antagonists
 CC may also be used as therapeutic agents to down regulate expression and
 CC activity. The antibodies may also be used as diagnostic agents for
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme
 CC linked immunosorbent assay (ELISA)). Examples of diseases which may be
 CC treated include rheumatoid arthritis and diabetes

SQ Sequence 2510 BP; 802 A; 412 C; 489 G; 807 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.51e-32 Length: 2510
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 5 Gaps: 0

US-10-028-384-8 (1-774) x AAF93772 (1-2510)

QY 551 AlaArgValMetSerTrpTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
 Db 268 GCACGAGTAATGCTTGGTGGATATGGCTATCAGATAGCTGAATGGCTAATAGAACT 327

QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 Db 328 ACGTTGGTGGATAAATACACCTGGATAACAGCCACATAGCACTGGTGGAAAGCTATG 387

QY 591 SerSer 592
 Db 388 TCTTCT 393

RESULT 14

AAD08289

ID AAD08289 standard; cDNA; 2537 BP.

XX AC AAD08289;

XX DT 08-AUG-2001 (first entry)

XX DE Human secreted protein-encoding gene 7 cDNA clone HDTLR06, SEQ ID NO: 17.
 XX KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;
 KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
 KW cell culture; chemotaxis; vulnerability; binding partner identification;
 KW gene therapy; ss.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers

FT CDS 93..953

FT /tag= a

FT /product= "Human secreted protein precursor"

FT sig_peptide 93..155

FT /tag= b

FT mat_peptide 156..950

FT /tag= c

FT /product= "Mature human secreted protein"

XX WO200136440-A1.

XX 25-MAY-2001.

XX

us-10-028-384-8.Oligo.rng

Thu Dec 16 16:25:14 2004

PF 15-NOV-2000; 2000WO-US0311282.
XX
PR 19-NOV-1999; 99US-0166414P.
PR 21-JUL-2000; 2000US-0219665P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsculis GA, Birse CE, Moore PA;
XX
XX WPI: 2001-343795/36.
DR P-PSDB; AAE03824.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating a medical condition.
XX
XX Claim 1; Page 443; 553pp; English.
XX
XX AAD08283-AAD08355 represent cDNAs corresponding to 23 human secreted
CC protein genes, and AAE03818-AAE03870 represent the proteins they encode.
CC AAE03871-AAE03896 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 23 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g. rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g. Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
CC present sequence represents a human secreted protein-encoding cDNA of the
CC invention
XX
XX SQ Sequence 2537 BP; 816 A; 434 C; 487 G; 800 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1-52e-32 Length: 2537
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x AAD08289 (1-2537)
QY 551 AlahrgValMetSerTrpTppAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
DB 267 GCACGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGATGGCTATAGACT 326
QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
DB 327 ACGTTGGTGGATATAACACCTGGGATTAACAGCCACATAGCACTGGTGGGAAAGCTATG 386
QY 591 SerSer 592
DB 387 TCCTCT 392
RESULT 15
AAV44866

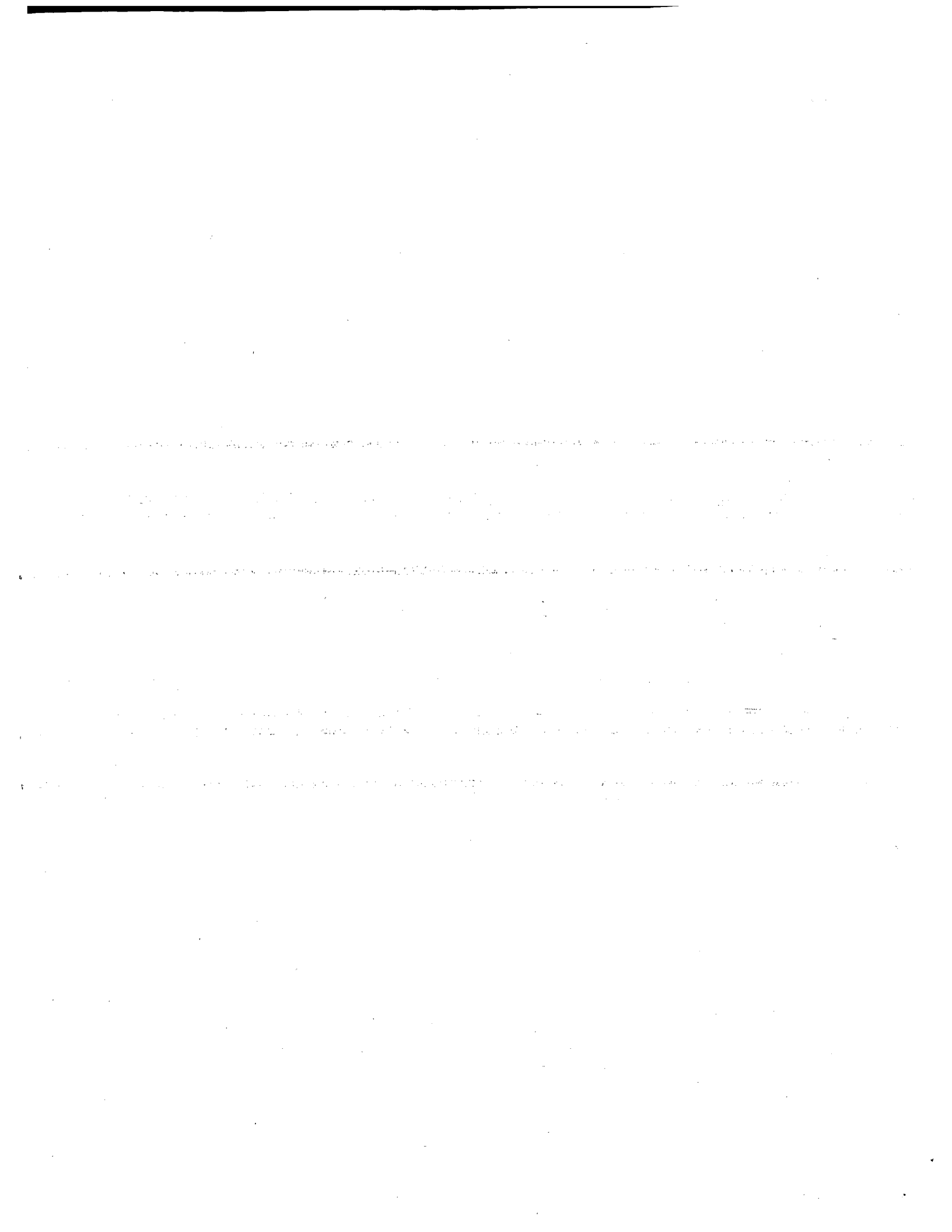
AAV44866 standard; cDNA; 2546 BP.
AAV44866;
21-OCT-1998 (first entry)
Clone CT585_1 coding sequence.
Secreted protein; nutritional source; cell proliferation activity;
cell differentiation activity; immune stimulant; tissue growth activator;
haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor;
tumour inhibitor; clone CT585_1; ds.
Homo sapiens.
Homo sapiens.
Key Location/Qualifiers
CDS 112..972
/*tag= a
WO9825962-A2.
18-JUN-1998.
12-DEC-1997; 97WO-US0232224.
13-DEC-1996; 96US-00766263.
11-DEC-1997; 97US-00989232.
(GEMY) GENETICS INST INC.
Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
Spaulding V, Agostino WJ;
WPI: 1998-362424/31.
P-PSDB; AAW69247.
New isolated polynucleotides - obtained from human adult testis, human
adult ovary, human adult brain and human adult heart cDNA libraries.
Claim 35; Page 79-81; 108pp; English.
This sequence represents a polynucleotide of the invention, and encodes a
secreted protein. It was isolated from a human adult brain cDNA library,
and is designated clone CT585_1. The DNA sequences and encoded
polypeptides can be used as nutritional sources or supplements, or may
exhibit e.g. cytokine and cell proliferation/differentiation activity,
immune stimulating or suppressing activity, haematopoiesis regulating
activity, receptor/ligand activity, anti-inflammatory activity,
activin/inhibin activity, chemostatic/chemokinetic activity,
cadherin/tumour invasion suppressor activity, tissue growth activity,
tumour inhibition activity or other activities
SQ Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1-53e-32 Length: 2546
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 2 Gaps: 0

US-10-028-384-8 (1-774) x AAV44866 (1-2546)
QY 551 AlahrgValMetSerTrpTppAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
DB 286 GCACGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGATGGCTATAGACT 345
QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
DB 346 ACGTTGGTGGATATAACACCTGGGATTAACAGCCACATAGCACTGGTGGGAAAGCTATG 405
QY 591 SerSer 592

Db |||||
 406 TCTTCT 411

Search completed: December 14, 2004, 23:35:02
Job time : 918.777 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 23:35:31 ; Search time 974.477 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 774

Sequence: 1 MNRTPMLNSKVGYSLLT.....RRKGYIENRPPVVGKRTLK 774

Scoring table:

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8161859

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat.14122004.131648.6892/app_query.fasta.1.3740
-DB=Published Applications NA -QFWT=fastap -SURFIX=Oligo.rnpb -MINMATCH=0.1
-LOPCU=0 -LOPEXI=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THRS=ext -HEAPSIZE=500 -MINLEN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -H2A24 @runat.14122004.131648.6892
-MAXLEN=2000000000 -USER=US10028384 @cgn.1.1.2124 @runat.14122004.131648.6892
-NCPU=6 -ICPU=3 -NO_MMAP -LARGESOURCY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09C_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	774	100.0	2417	15	US-10-028-384-7
2	42	5.4	430	9	US-09-878-178-717
3	42	5.4	430	13	US-10-046-935-717
4	42	5.4	430	14	US-10-146-502-717
5	42	5.4	1828	10	US-09-945-527-62
6	42	5.4	2481	15	US-10-028-384-7
7	42	5.4	2481	15	US-10-028-384-7
8	42	5.4	2110	15	US-10-028-384-3
9	34	4.4	387	16	US-10-276-774-173
10	34	4.4	558	14	US-10-052-283-433
11	34	4.4	1209	15	US-10-106-698-330
12	34	4.4	1209	16	US-10-264-237-412
13	34	4.4	1543	10	US-09-974-879-133
14	34	4.4	1543	10	US-09-305-736-133
15	34	4.4	1543	10	US-09-818-683-133
16	34	4.4	1543	11	US-09-818-683-133
17	34	4.4	1543	16	US-10-621-401-133
18	32	4.1	1848	15	US-10-128-714-1139
19	32	4.1	1969	15	US-10-128-714-1139
20	32	4.1	2232	15	US-10-128-714-1139
21	32	4.1	2603	15	US-10-128-714-1139
22	32	4.1	3969	15	US-10-128-714-1139
23	32	4.1	4603	15	US-10-128-714-1139
24	30	3.9	743	10	US-09-945-527-63
25	30	3.9	1114	16	US-10-296-115-629
26	28	3.6	500	9	US-09-998-598-1643
27	28	3.6	616	9	US-09-879-536-332
28	27	3.5	474	14	US-10-052-283-454
29	23	3.0	637	15	US-10-002-631C-183
30	23	3.0	640	15	US-10-002-631C-179
31	21	2.7	154	10	US-09-991-936-450
32	21	2.7	2472	15	US-10-171-581-112
33	21	2.7	2472	15	US-10-028-384-11
34	21	2.7	2472	15	US-10-172-118-742
35	21	2.7	2472	16	US-10-342-887-742
36	21	2.7	3046	18	US-10-417-375-95
37	21	2.7	3093	18	US-10-417-375-92
38	21	2.7	3094	15	US-10-028-384-9
39	21	2.7	5404	18	US-10-417-375-99
40	21	2.7	5827	18	US-10-417-375-97
41	20	2.6	393	16	US-10-276-774-123
42	20	2.6	502	17	US-10-767-701-6513
43	20	2.6	2157	9	US-09-801-368-387
44	20	2.6	2157	18	US-10-793-639-318
45	20	2.6	2681	16	US-10-425-114-14408

ALIGNMENTS

RESULT 1

US-10-028-384-7
Sequence 7, Application US/10028384
Publication No. US20030148285A1
GENERAL INFORMATION:
APPLICANT: COMPATIGENE INC.
APPLICANT: PERREAU, Claude
APPLICANT: MCBRIDE, Kevin
TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
FILE REFERENCE: 5600-74
CURRENT APPLICATION NUMBER: US/10/028.384
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patent in version 3.1
SEQ ID NO 7
LENGTH: 2417
TYPE: DNA
ORGANISM: Drosophila melanogaster
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: AF132552
DATABASE ENTRY DATE: 1999-04-27

; RELEVANT RESIDUES: (1)..(2417)

US-10-028-384-7

Alignment Scores:

Pred. No.: 0
 Score: 774.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 DB: 15
 Gaps: 0

US-10-028-384-8 (1-774) x US-10-028-384-7 (1-2417)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20
 DB 71 ATGATCGGAGCCGAGAGATGCTGAACAGCAAGTGGCTGCTACAGAGCTTAATCACC 130
 QY 21 PheAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40
 DB 131 TTCGCCATCCTGCTAATCGCTGGCTGGCCGGATTTTCCTCTCGCCCTCTCGCCGTCATC 190
 QY 41 ArgPheGluSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 60
 DB 191 CGTTTCGAGTCGATATCCATGAGTTTGATCCGCTGGTTCAACTACCGGGCCACCCCTAC 250
 QY 61 MetValGlnAsnGlyTyrPheAsnPheLeuAsnTyrPheAspGluArgAlaThrLeuTyr 80
 DB 251 ATGGTGAGAGATGGTTGGTCAACTCTCCTCAACTGGTTTCGACGAGCGGCATGGTATCGG 310
 QY 81 LeuGlyArgLeuValGlyThrValTyrProGlyLeuMetLeuLeuLeuLeuLeuLeuLeu 100
 DB 311 CTCGGCAGGATTTGGCGGCTACCGTCTATCCCGGCTGATGATTACGTCGGCGGAATC 370
 QY 101 HisTrpLeuLeuHisValLeuAsnLeuProValHisLeuLeuLeuLeuLeuLeuLeuLeu 120
 DB 371 CATTTGGCTGTGACGTAACATACACGCTGCTATTCGATGACATCTGCTGCTGCTGCTG 430
 QY 121 AlaProLeuPheSerGlyLeuThrSerLeuThrTyrLeuLeuLeuLeuLeuLeuLeuLeu 140
 DB 431 CGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACTCTGCTGACNAGGAGGTGG 490
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheLeuLeuLeuLeuLeuLeuLeuLeuLeu 160
 DB 491 TCCGCGGGCGCGGCTCTTCGCGCGAGCTTCATCGCATCTGCTGCTGCTGCTGCTGCTG 550
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyLeuAlaLeuPheAlaLeuGlnPheThr 180
 DB 551 AGGTGGTGGCTGGATCGTACGATACGAGGGGATTTGCCATATTGCGCTGCGATTCACC 610
 QY 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAla 200
 DB 611 TACTTCTGTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 670
 QY 201 LeuSerTyrPheTyrMetValSerAlaTyrGlyTyrValPheLeuLeuLeuLeuLeuLeu 220
 DB 671 TTGTCTTCTTCTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 730
 QY 221 ProLeuHisValPheValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 240
 DB 731 CCCCTGACGCTCTTCGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 790
 QY 241 TyrSerThrPheTyrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 260
 DB 791 TACAGCACCTTCTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850
 QY 261 GlnProLeuArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 DB 851 CAACCGATACGACCGACGATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 910
 QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 DB 911 GTGCCACCTTTGGCCATTGGAGTCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 970

QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValValLeuThrMet 320
 DB 971 ATCGTCGGCGGATTTGCTGGCGCTTGTGGCGCTTGTGGCGCTTGTGGCGCTTGTGGCGCT 1030
 QY 321 LeuGlyValValAlaProTyrPheSerGlyArgPheTyrSerLeuTyrAspThrGlyTyrAla 340
 DB 1031 CTGGGGGTTGTGGCCCGCTGGAGTGGAGCGCTTCTACTGCTGTGGGATAGTGGTAGGCC 1090
 QY 341 LysIleHisLeuProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 360
 DB 1091 AAGATCCACATCCCATCATTCGATCCGCTGGAGCATCAGCCACCATTCGCTGCTTCTCG 1150
 QY 361 PhePhePheAspLeuHisLeuValCysAlaPheProValGlyValTyrTyrCysIle 380
 DB 1151 TTCTTTTGTGATCTGCACATCTGCTGGCTTCCAGCTGGAGTGGAGTGGAGTGGAGTGGAG 1210
 QY 381 LysGlnLeuAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 DB 1211 AAGCAGATCAACGACGAGCGCGCTTTTCTGCTGGTGTGTACGCCATCAGTCGCGTTTACTTC 1270
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 DB 1271 GCTGGTGTGATGGTGGCTTGTGATGGTGTGATGGTGTGATGGTGTGATGGTGTGATGG 1330
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnLeuAspSerSerLysArgMetGly 440
 DB 1331 GTGGCCCTTTTCGGGACTGTTGGATGTTCTCTGCAAGAGGATTCGTTCTAAGCGAATGGC 1390
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerLeuGluLysLysThr 460
 DB 1391 ACAGCCATTAAGCGAGCCACCGAAGTGGATGAGCTGAGGATTCATTTAGAGAGAACG 1450
 QY 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThr 480
 DB 1451 CTGTACGACAAAGCTGGCAAGCTGAAGCATGCTGTAAGCATGATGATGATGATGATGATG 1510
 QY 481 GlyValSerSerAsnLeuLysSerLeuValLeuAlaValLeuMetLeuLeuMetMet 500
 DB 1511 GCGCTCAGCTCCAACTGAGAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1570
 QY 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerLeuValLeu 520
 DB 1571 TTGCTGTCTCACTGACGCTGGTGACCAAGCAATGCTACTCCAGTCCCTCCATTTGCTT 1630
 QY 521 AlaPheHisAsnSerGlnAspGlySerArgAsnLeuLeuAspAspPheArgGluAlaTyr 540
 DB 1631 GCTTTCCACACAGTCAAGATGGATCCCGCAACATTTTAGACGATTTACAGAGGCTTAC 1690
 QY 541 TyrTrpLeuSerGlnAsnThrAlaAspAlaArgValMetSerTrpTyrAspTyrGly 560
 DB 1691 TACTGGCTTTCGACAGACACTGCGGATGATGCTGCGCTTATGCTTGGTGGGATTACGGA 1750
 QY 561 TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnThrTrpAsnAsn 580
 DB 1751 TACCAGATAGCGGATGGCAACAGACGAGCTAGTGGNATATATATAGTGGNACAT 1810
 QY 581 SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle 600
 DB 1811 AGTCATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGATCTCTACGAAAT 1870
 QY 601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620
 DB 1871 ATGACATCTTACGCTGGACTAGCTTTGGTGTATCTTGGCGGTGATCGGCTATTCT 1930
 QY 621 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 640
 DB 1931 GCGCATGATATCAACAAGTCTCTGAGTGTCTGCAATTCGAGGAGAGATCCCAAG 1990
 QY 641 AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly 660
 DB 1991 GACATTAAGGAAGCGATTACTTTTACCGACCGCGTGAATTCAGGCTAGATGCCAAGGT 2050
 QY 661 AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680

Db 2051 GCTCGGCGCTGCTCACTGCTTATGACAAATTAAGCTACTACAGATTCGGGGAATG 2110
QY 681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
Db 2111 AAGTTGGACTACAGAGGTCCATCTGGATATGATCCACAGCAAGCGGTCAATGGGAAT 2170
QY 701 LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrGluHisTyrPleuValArg 720
Db 2171 AAGGACTTCGATCTGACCTACTGGAGGAGGCGCTACACAGACACTGCTTGTTCGC 2230
QY 721 IleTyrArgValLysPheProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
Db 2231 ATCTATAGGTGAAGAGCGCATGAGTTCAATAGACATCACTGGAAGACCAAGAGGAGA 2290
QY 741 ThrIleProProAlaAsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIle 760
Db 2291 AGATTCTCCAGCAAACTTCATTTCGAGAAAGAACTCTAAGCGCTCCAGGGCTACATA 2350
QY 761 ArgAsnArgProValValValLysGlyLysArgThrLeuLys 774
Db 2351 CGAAACCGACCGTGTGTGTTAAGGGAACCAACCTTGAAA 2392

RESULT 2

US-09-878-178-717
; Sequence 717, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-878-178-717
Alignment Scores:
Pred. No.: 7,59e-34 Length: 430
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 9 Gaps: 0
US-10-028-384-8 (1-774) x US-09-878-178-717 (1-430)
QY 551 AlaArgValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
Db 26 GCACAGTAATGCTCTGTGGATATGCTATCAGATAGCTGAATGGCTAATAGAACT 85
QY 571 ThrLeuValAspAsnAsnThrTyrAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
Db 86 ACGTTGGTGGATAATAACACCTGGATAACAGCCACATACACTGGTGGGAAAAGCTATG 145
QY 591 SerSer 592
Db 146 TCTTCT 151

RESULT 4

US-10-146-502-717
; Sequence 717, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secretist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-146-502-717
Alignment Scores:
Pred. No.: 7,59e-34 Length: 430
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 14 Gaps: 0
US-10-028-384-8 (1-774) x US-10-146-502-717 (1-430)
QY 551 AlaArgValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

RESULT 3

US-10-046-935-717
; Sequence 717, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.

```

Db      26 GCACGAGTAATGCTTGGTGGATTATGGCTATCAGATAGCTGGTAATGGCTAATAGAACT 85
QY      571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
        |||||
Db      86 ACGTTGGTGGATAATAACACCTGGATAACAGCCACATAGCACTGGTGGGAAAAGCTATG 145
QY      591 SerSer 592
        |||||
Db      146 TCTTCT 151

RESULT 5
US-09-945-527-62
; Sequence 62, Application US/09945527
; Publication No. US20030055588A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20030055588A1el Nucleic Acid Molecules Encoding
; FILE REFERENCE: Nucleic Acid and Protein Homologs
; CURRENT FILING DATE: 2001-08-29
; CURRENT APPLICATION NUMBER: US/09/945,527
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 1828
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1767..1771
; OTHER INFORMATION: n = A, T, C or G
US-09-945-527-62

Alignment Scores:
Pred. No.: 2,96e-33 Length: 1828
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x US-09-945-527-62 (1-1828)
QY      551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
        |||||
Db      451 GCACGAGTAATGCTTGGTGGATTATGGCTATCAGATAGCTGGTAATGGCTAATAGAACT 510
QY      571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
        |||||
Db      511 ACGTTGGTGGATAATAACACCTGGATAACAGCCACATAGCACTGGTGGGAAAAGCTATG 570
QY      591 SerSer 592
        |||||
Db      571 TCTTCT 576

RESULT 6
US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT FILING DATE: 2001-12-20
; CURRENT APPLICATION NUMBER: US/10/028,384
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)
; OTHER INFORMATION:
US-10-028-384-1

Alignment Scores:
Pred. No.: 3,94e-33 Length: 2481
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-8 (1-774) x US-10-028-384-1 (1-2481)
QY      551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
        |||||
Db      1795 GCACGAGTAATGCTTGGTGGATTATGGCTATCAGATAGCTGGTAATGGCTAATAGAACT 1854
QY      571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
        |||||
Db      1855 ACGTTGGTGGATAATAACACCTGGATAACAGCCACATAGCACTGGTGGGAAAAGCTATG 1914
QY      591 SerSer 592
        |||||
Db      1915 TCTTCT 1920

RESULT 7
US-10-264-049-630
; Sequence 630, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 630
; LENGTH: 2660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2634)..(2634)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2637)..(2637)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-630

Alignment Scores:
Pred. No.: 4,21e-33 Length: 2660
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 16 Gaps: 0

US-10-028-384-8 (1-774) x US-10-264-049-630 (1-2660)
QY      551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
        |||||
Db      330 GCACGAGTAATGCTTGGTGGATTATGGCTATCAGATAGCTGGTAATGGCTAATAGAACT 389
QY      571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
        |||||

```

Db 390 ACCTGGTGGTAAATACACCTTGGTAATACACCCACATGACCTGGTGGGAAGCTATG 449
QY 591 SerSer 592
Db 450 TCTTCT 455
RESULT 8
US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAULT, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)...(2469)
US-10-028-384-3
Alignment Scores:
Pred. No.: 4,29e-33 Length: 2710
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 15 Gaps: 0
US-10-028-384-8 (1-774) x US-10-028-384-3 (1-2710)
QY 551 AlaArgValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
Db 1862 GCCCGGGTCATGTCGTGGTGGACTACGGCTATCAGATGCTGGCATGCCAACAGNCC 1921
QY 571 ThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyIleAlaMet 590
Db 1922 ACTCTGGTGGATAACAACACCTGGACAAACAGCCACATGGCACTGGTGGAAAAAGCTATG 1981
QY 591 SerSer 592
Db 1982 TCTTCC 1987
RESULT 9
US-10-276-774-173
; Sequence 173, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 173
; LENGTH: 387
; TYPE: DNA

; ORGANISM: Homo sapiens
US-10-276-774-173
Alignment Scores:
Pred. No.: 1.62e-25 Length: 387
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.39% Indels: 0
DB: 16 Gaps: 0
US-10-028-384-8 (1-774) x US-10-276-774-173 (1-387)
QY 325 AlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIle 344
Db 235 GCACCATGGAGTGGCAGGTTTTATTTCATGTGGGATACCTGGGTATGCAAAAATACACATT 294
QY 345 ProIleIleAlaSerValSerGluHisGlnProThrThrTrp 358
Db 295 CCAATTATTGCATCAGTGTCTGAGCAATCAACCTACGACTTGG 336
RESULT 10
US-10-052-283-433
; Sequence 433, Application US/10052283
; Publication No. US20030064379A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Goddard, Audrey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND METHOD OF USE THEREOF
; FILE REFERENCE: P2751R1C1
; CURRENT APPLICATION NUMBER: US/10/052,283
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: PCT/US00/20006
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/145,701
; PRIOR FILING DATE: 1999-07-26
; NUMBER OF SEQ ID NOS: 564
; SEQ ID NO 433
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 23, 111, 130, 146
; OTHER INFORMATION: unknown base
US-10-052-283-433
Alignment Scores:
Pred. No.: 2.29e-25 Length: 558
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.39% Indels: 0
DB: 14 Gaps: 0
US-10-028-384-8 (1-774) x US-10-052-283-433 (1-558)
QY 325 AlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIle 344
Db 224 GCACCATGGAGTGGCAGGTTTTATTTCATGTGGGATACCTGGGTATGCAAAAATACACATT 283
QY 345 ProIleIleAlaSerValSerGluHisGlnProThrThrTrp 358
Db 284 CCAATTATTGCATCAGTGTCTGAGCAATCAACCTACGACTTGG 325
RESULT 11
US-10-106-698-330
; Sequence 330, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide

FILE REFERENCE: PA005P1
 CURRENT APPLICATION NUMBER: US/10/106,698
 CURRENT FILING DATE: 2002-03-27
 PRIOR APPLICATION NUMBER: PCT/US00/26524
 PRIOR FILING DATE: 2000-09-28
 PRIOR APPLICATION NUMBER: US 60/157,137
 PRIOR FILING DATE: 1999-09-29
 PRIOR APPLICATION NUMBER: US 60/163,280
 PRIOR FILING DATE: 1999-11-03
 NUMBER OF SEQ ID NOS: 8564
 SOFTWARE: PatentIn Ver. 3.0
 SEQ ID NO 330
 LENGTH: 1209
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (1203)..(1203)
 OTHER INFORMATION: n equals a,t,g, or c
 US-10-106-698-330

Alignment Scores:
 Pred. No.: Length: 1209
 Score: 4.73e-25
 Matches: 34
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 4.39%
 Indels: 0
 Gaps: 0
 DB: 0

US-10-028-384-8 (1-774) x US-10-106-698-330 (1-1209)

Qy 325 AlaProTrrSerGlyArgPheTySerLeuTrpAspThrGlyTyrAlaLysIleHisIle 344
 Db 866 GCACCATGGAGTGGCAGGTTTTTATTCAATTGTGGGATCTGGGTATGCAAAAATACACATT 925
 Qy 345 ProIleLeAlaSerValSerGluHisGlnProThrTrp 358
 Db 926 CCAATTATTGTCATCAGTGTCTGAGCATCAACCTACGACTTGG 967

RESULT 12

US-10-264-237-412
 Sequence 412, Application US/10254237
 Publication No. US20040009491A1
 GENERAL INFORMATION:
 APPLICANT: Birse et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 FILE REFERENCE: PA131P1
 CURRENT APPLICATION NUMBER: US/10/264,237
 CURRENT FILING DATE: 2002-10-04
 PRIOR APPLICATION NUMBER: PCT/US01/16450
 PRIOR FILING DATE: 2001-05-18
 PRIOR APPLICATION NUMBER: US 60/205,515
 PRIOR FILING DATE: 2000-05-19
 NUMBER OF SEQ ID NOS: 2876
 SOFTWARE: PatentIn Ver. 3.1
 SEQ ID NO 412
 LENGTH: 1209
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (1203)..(1203)
 OTHER INFORMATION: n equals a,t,g, or c
 US-10-264-237-412

Alignment Scores:
 Pred. No.: Length: 1209
 Score: 4.73e-25
 Matches: 34
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 4.39%
 Indels: 0
 Gaps: 0
 DB: 0

US-10-028-384-8 (1-774) x US-10-264-237-412 (1-1209)

Qy 325 AlaProTrrSerGlyArgPheTySerLeuTrpAspThrGlyTyrAlaLysIleHisIle 344
 Db 866 GCACCATGGAGTGGCAGGTTTTTATTCAATTGTGGGATCTGGGTATGCAAAAATACACATT 925
 Qy 345 ProIleLeAlaSerValSerGluHisGlnProThrTrp 358
 Db 926 CCAATTATTGTCATCAGTGTCTGAGCATCAACCTACGACTTGG 967

RESULT 13

US-09-974-879-133
 Sequence 133, Application US/09974879
 Publication No. US20030028003A1
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: 125 Human Secreted Proteins
 FILE REFERENCE: P202P2
 CURRENT APPLICATION NUMBER: US/09/974,879
 CURRENT FILING DATE: 2001-10-12
 PRIOR APPLICATION NUMBER: US 60/239,893
 PRIOR FILING DATE: 2000-10-13
 PRIOR APPLICATION NUMBER: US 09/818,683
 PRIOR FILING DATE: 2001-03-28
 PRIOR APPLICATION NUMBER: US 09/305,736
 PRIOR FILING DATE: 1999-05-05
 PRIOR APPLICATION NUMBER: PCT/US98/23435
 PRIOR FILING DATE: 1998-11-04
 PRIOR APPLICATION NUMBER: US 60/064,911
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,912
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,983
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,900
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,988
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,987
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,909
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,984
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,985
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/066,094
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: US 60/066,100
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: US 60/066,089
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: US 60/066,095
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: US 60/066,090
 PRIOR FILING DATE: 1997-11-17
 NUMBER OF SEQ ID NOS: 611
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 133
 LENGTH: 1543
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (1055)
 OTHER INFORMATION: n equals a,t,g, or c
 NAME/KEY: SITE
 LOCATION: (1143)
 OTHER INFORMATION: n equals a,t,g, or c
 US-09-974-879-133

Alignment Scores:
 Pred. No.: Length: 1543
 Score: 5.95e-25

Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.39% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x US-09-974-879-133 (1-1543)

QY 325 AlaProTtpSerGlyArgPheTyrSerLeuTtpAspThrGlyTyrAlaLysIleHisIle 344
|||
Db 866 GCACCATGGAGTGGCAGGTTTATTTCATTGTTGGTATCTGGTATGCAAAATACACATT 925
|||
QY 345 ProIleIleAlaSerValSerGluHisGlnProThrThrTtp 358
|||
Db 926 CCAATTATTGCATCAGTGTCTGAGCATCAACCTACGACTTGG 967
|||

RESULT 14

US-09-305-736-133
; Sequence 133 Application US/09305736
; Publication No. US2003008078A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/305,736
; EARLIER FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: PCT/US98/23435
; EARLIER FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/064,911
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,912
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,983
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,900
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,988
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,987
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,908
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,984
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,985
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/066,094
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,100
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,089
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,095
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,090
; EARLIER FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-305-736-133

Alignment Scores:

Pred. No.: 5,95e-25 Length: 1543
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.39% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x US-09-305-736-133 (1-1543)

QY 325 AlaProTtpSerGlyArgPheTyrSerLeuTtpAspThrGlyTyrAlaLysIleHisIle 344
|||
Db 866 GCACCATGGAGTGGCAGGTTTATTTCATTGTTGGTATCTGGTATGCAAAATACACATT 925
|||
QY 345 ProIleIleAlaSerValSerGluHisGlnProThrThrTtp 358
|||
Db 926 CCAATTATTGCATCAGTGTCTGAGCATCAACCTACGACTTGG 967
|||

RESULT 15

US-09-818-683-133
; Sequence 133 Application US/09818683
; Publication No. US20030211472A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/818,683
; CURRENT FILING DATE: 2001-03-28
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-818-683-133

Alignment Scores:

Pred. No.: 5,95e-25 Length: 1543
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.39% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x US-09-818-683-133 (1-1543)

QY 325 AlaProTtpSerGlyArgPheTyrSerLeuTtpAspThrGlyTyrAlaLysIleHisIle 344
|||
Db 866 GCACCATGGAGTGGCAGGTTTATTTCATTGTTGGTATCTGGTATGCAAAATACACATT 925
|||
QY 345 ProIleIleAlaSerValSerGluHisGlnProThrThrTtp 358
|||
Db 926 CCAATTATTGCATCAGTGTCTGAGCATCAACCTACGACTTGG 967
|||

Search completed: December 15, 2004, 15:27:41
Job time : 993.477 secs

Blank
Page

QY 41 ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60
 DB 202 CGTTTCGAGTCGATATCCATGAGTTTGATCCGGTGGTTCAACTACCGGGCCCGGCTAC 261
 QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluAlaThrAlaTyrPro 80
 DB 262 ATGGTGCAGAAATGGTGGTCAACTTCCTCAACTGGTTCGACGAGCGCATGGTATCCG 321
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 DB 322 CTCGGCAGATTGTGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGGCGGAATC 381
 QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 DB 382 CATTGGCTGCTGCACGACTCAACATACCGGTCCATATTCGTGACATCTGCGTGTTCCTG 441
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 DB 442 GCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 501
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 DB 502 TCCGCGGCGCGCGCTCTTCGCGCGCAGCTTCATCGCCATCGCTGCTGGCTGACATCAGT 561
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 DB 562 AGGTCGGTGGCTGATCGTACGATAACGAGGCGATTGCCATATTCGCCCTGCAGTTCACC 621
 QY 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaIleAla 200
 DB 622 TACTTCCTGGTGGCGCTCAGTGAAGACTGGATCCGTGTCTGTGGTGGCGCGCGCGCT 681
 QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle 220
 DB 682 TTGCTCTACTTCTACATGGTGTCCGCTGGGTGGTACGTGTTCATCATCACTGATA 741
 QY 221 ProLeuHisValPheValLeuLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSer 240
 DB 742 CCCCTGCACGCTTCCTGACTGCTCATTTATGGGAGGACTCTCCGCGGCTCTGCTGACCCAGC 801
 QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 DB 802 TACAGCACCTTCTACATCTCGGACCTGCTGTCTCCATGCAGATCCCTTCGTGGGATTC 861
 QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGly 273
 DB 862 CAACCGATACGACACGATGAACATGGCTGGCGTGGGA 900

RESULT 2

US-09-270-767-158
 ; Sequence 158, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 158
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-158

Alignment Scores:

Pred. No.: 1,26e-241 Length: 1660
 Score: 246.00 Matches: 246
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.78% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-158 (1-1660)

QY 529 SerArgAsnIleIleuAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAla 548
 DB 788 TCCCGCAACATTTTAGACGATTCAGAGAGGCTTACTACTGGCTTTCGCAAAACACGTC 847
 QY 549 AspAlaAlaArgValMetSerTyrTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 568
 DB 848 GATGATGCTCGCGTTATGCTTGGTGGGATTACGGATACCGATAGCGGGAATGGCAAC 907
 QY 569 ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 588
 DB 908 AGAACACCGCTAGTGGATAAATACGTGGAAACAATAGTACATAGCGCTGGTGGCAAG 967
 QY 589 AlaMetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyr 608
 DB 968 GCAATGCTTCAACCGAGGAGAAGTCTTACGAAATTAATGACATCTCTTGACGTGGACTAC 1027
 QY 609 ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 628
 DB 1028 GTTTTGGTGAATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGATATCAACAAGTTCTCTG 1087
 QY 629 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
 DB 1088 TGGATGGTCCGAATTGTCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTT 1147
 QY 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
 DB 1148 ACCGACCGCGGTGAATTCAGGGTAGATCCGAAGGTGCTCCGGCCCTGCTCAACTGCGCTT 1207
 QY 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
 DB 1208 ATGTACAATTAAGCTACTTACAGATTTCGGGAATTTGAAGTTGGACTACAGAGGTCCATCT 1267
 QY 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
 DB 1268 GGATATGATCGCACACGTAACCGCTCATTTGGGAATTAAGGACTTCGATCTGACCTACCTG 1327
 QY 709 GluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHis 728
 DB 1328 GAGGAGCGCTACACACAGAACACTGGCTTGTTCGCATCTATAGGTGAAGAAGCGCAT 1387
 QY 729 GluPheAsnArgProSerLeuLysThrLysGluArgThrIleProProAlaAsnPheIle 748
 DB 1388 GAGTTCAATAGACCATCTACTGAAGACCAAGAGAGACGATTCCTCCAGCAAACTTCATT 1447
 QY 749 SerArgLysAsnSerLysArgArgLysGlyTyrIleArgAsnArgProValValLys 768
 DB 1448 TCGAGAAAGAACTTAAGCGCTCGCAAGGCTTACATACGAAACCGCGGTGTGTTTAAG 1507
 QY 769 GlyLysArgThrLeuLys 774
 DB 1508 GGAACCAAGCACTTGAAA 1525

RESULT 3

US-09-270-767-15440
 ; Sequence 15440, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15440
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-15440

```
Alignment Scores:
Pred. No.: 1.26e-241 Length: 1660
Score: 246.00 Matches: 246
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.78% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-15440 (1-1660)
QY 529 SerArgAsnIleLeuAspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAla 548
DB 788 TCCCGCAACATTTAGACGATTCAGAGAGCGCTTACTAGCTTTCGCGAAGACAGTGC 847
QY 549 AspAspAlaArgValMetSerTrpAspTyrGlyTrpGlnIleAlaGlyMetAlaAsn 568
DB 848 GATGATGCTCGGTTATGCTTTGGTGGATTCAGGATACCATAGGCGGAATGGCAAC 907
QY 569 ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisLeAlaLeuValGlyLys 588
DB 908 AGAAGCGAGCTAGTGGAATAATACGTGGGAACAATAGTCACATAGCGCTGGTTGCAAG 967
QY 589 AlaMetSerSerThrGluGlyLysSerTyrGluIleMetThrSerLeuAspValAspTyr 608
DB 968 GCAATGCTTCAACCGAGAGAGAGTCCCTACGAAATATGACATCTCTTGACGTGGACTAC 1027
QY 609 ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 628
DB 1028 GTTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGATATCAACAAGTTCCTG 1087
QY 629 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
DB 1088 TGATGGTCCGAATTCCTGAGGAGAGCATCCCAAGGACATTAAGAAACGCAATTACTTT 1147
QY 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
DB 1148 ACCGACCGCGTGAATTCAGGCTAGATCCGAGGTGCTCCGCGCTGCTCAACTGCCTT 1207
QY 669 MetTyrLysLeuSerTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
DB 1208 AGTACAAATTAAGCTACTCAGATTTCGGGAATTAAGTTGACTACAGAGTCCATCT 1267
QY 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
DB 1268 GGATATGATCGCACACGTAAACCGCTCATTTGGGAATAAGGACTTCGATCTGACCTACCTG 1327
QY 709 GluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHis 728
DB 1328 GAGGAGGCTTACACACAGAACACTGGCTTTGTCGATCTATAGGGTGAAGAGCCGCAT 1387
QY 729 GluPheAsnArgProSerLeuLysThrLysGluArgThrIleProProAlaAsnPheIle 748
DB 1388 GAGTTCAATAGACCATCACTGAAGACCAAGGAGAGACGATTCTCCAGCAAACTTCATT 1447
QY 749 SerArgLysAsnSerLysArgArgLysGlyTyrIleArgAsnArgProValValLys 768
DB 1448 TCGAGAAAGAACTTAAGCGTCGCAAGGGCTACATACAGAAACCGCGGTGTGTGTTAAG 1507
QY 769 GlyLysArgThrLeuLys 774
DB 1508 GGAAGAACGAACTTGAAA 1525

RESULT 4
US-09-270-767-1318/c
; Sequence 1318, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517

Alignment Scores:
Pred. No.: 3.42e-156 Length: 560
Score: 162.00 Matches: 162
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.93% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-1318 (1-560)
QY 369 ValCysAlaPheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgVal 388
DB 492 GTGTGCGCCTTCCAGTGGAGTGTGTACTGTATCAAGCAGATCAACGACGCGCTT 433
QY 389 PheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMet 408
DB 432 TTCGTGTGTGTGTACGCCATCAGTGGGTACTTTCGCTGGTGTGTGTGGTGTGATG 373
QY 409 LeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAsp 428
DB 372 TTGACCTCTACGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 313
QY 429 ValPheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGlu 448
DB 312 GTGTTCCTGACAGAGGATTCGTCTAGCGAATGGCACAGCCATAAGCGCAGCCACCGAA 253
QY 449 ValAspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeu 468
DB 252 GTGATGAGCTGAGGATTCATTTGAGAAGAGACGCTGTACCAAGGCTGCAAGCTG 193
QY 469 LysHisArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSer 488
DB 192 AAGCATCGTACTAAGCATGATGCCACAGCATACTGGCGTCAGCTCCACCTGAAGAGT 133
QY 489 IleValIleLeuAlaValLeuMetLeuLeuMetPheAlaValHisCysThrTrpVal 508
DB 132 ATTGTATTTTGGCGCTTCTAATGCTGTGTATGATGTGCTGTCTCCACTGCAGTGGGTG 73
QY 509 ThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly 528
DB 72 ACCAGCAATGCTACTCCAGTCCCTCCATTTGCTTGGCTTCCACACAGCTCAAGATGA 13
QY 529 SerArg 530
DB 12 TCCCGC 7

RESULT 5
US-09-270-767-16600/c
; Sequence 16600, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 16600
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16600

Alignment Scores:
Pred. No.: 3.42e-156 Length: 560
Score: 162.00 Matches: 162
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.93% Indels: 0
DB: 4 Gaps: 0
```

Score: 162.00 Matches: 162
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.93% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-15600 (1-560)

QY 369 ValCysAlaPheProValGlyValTrpTyrCysIleIysGlnIleAsnAspGluArgVal 388
 DB 492 GTGTGGCGCTCCAGTGGAGTGTGTACTGATCAGCAGATCAACGACGAGCGGT 433
 QY 389 PheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMet 408
 DB 432 TTCGTGTGTGTACGGCCATCAGTGCCTTTACTTCGCTGTGTGTGTGTGTGTGT 373
 QY 409 LeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAsp 428
 DB 372 TTGACCTTCACCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 313
 QY 429 ValPheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGlu 448
 DB 312 GTGTTCCTGCAAGAGATTCGCTTAAGCGAATGGGCACACGATTAAGCGACCGAA 253
 QY 449 ValAspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeu 468
 DB 252 GTGATGAGCTGAGGATTCATTTGAGAGAGAGACCTGTACGACAGGCTGGCAGCTG 193
 QY 469 LysHisArgThrLysHisAspAlaGlnInAspThrGlyValSerSerAsnLeuLysSer 488
 DB 192 AAGCATCGTACTAAGCATGATGCGCCAGCAGGATCTGGCTCAGCTCCCAACCTGAAGAT 133
 QY 489 IleValIleLeuAlaValLeuMetLeuLeuMetPheAlaValHisCysThrTrpVal 508
 DB 132 ATTGTATTGGCGGTTCTAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 73
 QY 509 ThrSerAsnAlaTyrSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly 528
 DB 72 ACCAGCAATGCTACTCAGTCCCTCATCTGTCTGTCTGTCTGTCTGTCTGTCTGT 13
 QY 529 SerArg 530
 DB 12 TCCCGC 7

RESULT 6

US-09-270-767-158/c
 ; Sequence 158, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 158
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-158

Alignment Scores:
 Pred. No.: 1,15e-152 Length: 1660
 Score: 159.00 Matches: 210
 Percent Similarity: 99.06% Conservative: 0
 Best Local Similarity: 99.06% Mismatches: 1
 Query Match: 20.54% Indels: 2
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-158 (1-1660)

QY 564 AlaGlyMetAlaAsnArgThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 583

DB 793 GCGGGATGCGAACAAGACGACGCTAGTGGATATATATAGCTGGACAATAGTCACATA 734
 QY 584 AlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIleMetThrSer 603
 DB 733 GCGCTGTTGGCAAGGCAATGTCTTCAACCGAGGAGAGAGTCTCTACGAAATATGACATCT 674
 QY 604 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 623
 DB 673 CTTGACGTGATACGTTTGGTGTATCTTTGGCGGTGTATCGGCTATTTCGGGCGATGAT 614
 QY 624 IleAsnLysPheLeuTrpMetValArgIleAlaGluGlyHisProLysAspIleLys 643
 DB 613 ATCAACAAGTCTCTGATGATGTCGGAATTCGCGAGGAGAGCATCCCAAGGACATTAAG 554
 QY 644 GluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAla 663
 DB 553 GAAAGCGATATCTTTACCGACCGCGTGATTTACGGGTAGATGCGGAAGGTGCTCCGGCC 494
 QY 664 LeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAsp 683
 DB 493 CTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGGAATTGAGTTGGAC 434
 QY 684 TyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
 DB 433 TACAGAGGCGCATCTGGATATGTCGACACGTAACGCGTCATTTGGGAATTAAGSACTTT 374
 QY 704 AspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArg 723
 DB 373 GATCTGACCTACCTCGAGGAGGCGCTACACACAGAACACTGGCTTGTTCGCATCTATAG- 315
 QY 724 -VallyLysProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIlePr 743
 DB 314 TGTGAAGAAGCGCATGATTCATAGACCATCACTGAGACCCAGGAGAGACCATTC 255
 QY 743 oProAlaAsnPhelIleSerArgLysAsnSerLysArgLysGlyTyrIleArgAsnAr 763
 DB 254 TCCAGCAAACTTCATCTCGAAGAAAGACTCGAAGCGTCCGAAGGCTACATACGAAACCG 195
 QY 763 gProValValLysGlyLysArgThrLeuLys 774
 DB 194 ACCGTTGTGTTAAGGGAACGACCTTGAA 161

RESULT 7

US-09-270-767-15440/c
 ; Sequence 15440, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15440
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-15440

Alignment Scores:
 Pred. No.: 1,15e-152 Length: 1660
 Score: 159.00 Matches: 210
 Percent Similarity: 99.06% Conservative: 0
 Best Local Similarity: 99.06% Mismatches: 1
 Query Match: 20.54% Indels: 2
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-15440 (1-1660)

QY 564 AlaGlyMetAlaAsnArgThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 583

Db 793 GCGGGAATGGCAACAGACGCTAGTGTGATATAATACGTGGAAACATAGTACATA 734
QY 584 AlalaValGlyLysAlaValSerThrGluGluLysSerThrGluLeuMetThrSer 603
Db 733 GCGCTGGTGGCAAGCAATGTCTTCAACCGAGGAGAGTCTTACGAAATTTATGACATCT 674
QY 604 LeuAspValAspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAsp 623
Db 673 CTGACGTGGACTAGCTTTTGGTGTATCTTGGCGGTGATCGGCTATTCGGCGATGAT 614
QY 624 IleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIleLys 643
Db 613 ATCAACAAGTTCCTGTGATGTCGAATTCGCCGAGGAGAGCATCCCAAGGACATTAG 554
QY 644 GluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAla 663
Db 553 GAAGCGATTAATCTTACCAGCCGGGTGAAATCAGGGTAGATGCGGAGGTGCTCCGGCC 494
QY 664 LeuLeuAsnLysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAsp 683
Db 493 CTGCTCAACTGCCTTATGTACAATTAAGCTACTACAGATTCGGGGATTTGAAGTTGGAC 434
QY 684 TyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
Db 433 TACAGAGCCCATCTGGATATGATCGCACACGTAAACCGCGTCAATGGGAATAAGGACTTT 374
QY 704 AspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArg 723
Db 373 GATCTGACCTACCTGGAGGAGGCTTACACAGAACACTGGCTTGTTCGCATCTATAG- 315
QY 724 -ValLysIleProHisGluPheAsnArgProSerLeuLysThrIleGluArgThrIlePr 743
Db 314 TGTGAAGAAGCCGATGATGATTCATACACCATCACTGAGACCAAGAGAGACGATTCC 255
QY 743 oProAlaAsnPheIleSerArgLysAsnSerLysArgGlyTyrIleArgAsnAr 763
Db 254 TCCAGCAAACTTCTCTCGAAGAAAGAACTCGAAGCGTCGACGGGTACATACGAACCG 195
QY 763 gProValValLysGlyLysArgThrLeuLys 774
Db 194 ACCGGTTGTTGTAAGGGAACCAACCTTGAAA 161

RESULT 8

US-09-270-767-28082
; Sequence 28082, Application US/09270767
; Patent No. 6703491

GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28082
; LENGTH: 133
; TYPE: DNA
; ORGANISM: Drosophila melanogaster

US-09-270-767-28082

Alignment Scores:

Pred. No.: 3,85e-36 Length: 133
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.68% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-28082 (1-133)

QY 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeu 249
Db 2 ATGGGAGGAGTACTCGCCGCTGTGTCGACCACTAGCAGCACTTCTACATCTCTGGAGCTG 61

QY 250 LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
Db 62 CTGTTCTCCATGCAGATCCCTTCGTGGGATTCACACCGATAGCACACATG 121
QY 270 AlaAlaLeuGly 273
Db 122 GCTGCGCTGGGA 133

RESULT 9

US-09-513-999C-1965
; Sequence 1965, Application US/09513999C
; Patent No. 6783961

GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59, US2, REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm

; SEQ ID NO 1965

; LENGTH: 507

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 122..505

US-09-513-999C-1965

Alignment Scores:

Pred. No.: 1.57e-33 Length: 507
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-513-999C-1965 (1-507)

QY 551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
Db 113 GCACAGTAATGTCCTGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACT 172
QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
Db 173 ACGTGTGGTGGATAATAACACCTGGAATAACAGCCACATAGCAGCTGGTGGGAAAGCTATG 232

QY 591 SerSer 592

Db 233 TCTTCT 238

RESULT 10

US-09-328-111-332
; Sequence 332, Application US/09328111
; Patent No. 6262333

GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.

US-10-028-384-8 (1-774) X US-09-614-221A-318 (1-2157)

QY 151 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 170
DB 445 TTTATAGCCATTGTCCTCCGGTATATATCTAGATCAGTGGCGGGTCTCTACGATAATGAG 504

RESULT 14

US-09-270-767-27262
; Sequence 27262, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27262
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-27262

Alignment Scores:
Pred. No.: 5,12e-10 Length: 503
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-27262 (1-503)

QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePhe 175
DB 2 GGGTATATCTCGATCGTGGCGGATCGTACGACAAATGAAGGCAATCGGCATTTC 58

RESULT 15

US-09-248-796A-3089
; Sequence 3089, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3089
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (25)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-248-796A-3089

Alignment Scores:
Pred. No.: 8.68e-10 Length: 867
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-248-796A-3089 (1-867)

QY 618 GlyTyrSerGlyAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGly 636

Db 433 GGTATTCTGGTGGATATTAAACAATTTCTTATGGATGGTAAGAATTGCTGAAGGT 489

Search completed: December 14, 2004, 23:43:55
Job time : 161.806 secs

Blank
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:58 : Search time 5991.52 Seconds
(without alignments)
4707.376 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 774

Sequence: 1 MNRTFMLNLSKVAGYSSLT.....RRKGYYRNPVWVGKRTLK 774

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgr2_1/USFTO.spool/US10028384/runat_14122004_131646_6807/app_query.fasta_1.3740
-DB=EST -Qfmt=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPEL=0 -DOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US10028384 @CGN 1 13960 @runat_14122004_131646_6807 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:
1: gb_est1:
2: gb_est2:
3: gb_est3:
4: gb_est4:
5: gb_est5:
6: gb_est6:
7: gb_est7:
8: gb_gss1:
9: gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	245	31.7	810	2	BF502026
2	221	28.6	701	2	BF503420
3	214	27.6	655	2	BF496296
4	214	27.6	715	7	CK657451
5	208	26.9	679	7	CK657649
6	208	26.9	681	7	CK659033
7	197	25.5	652	4	BI171940
8	196	25.3	666	4	BI357074
9	196	25.3	668	4	BI227902

10	195	25.2	641	1	AI257750
11	195	25.2	663	4	BG641172
12	194	25.1	616	2	BF504378
13	191	24.7	630	1	AI295381
14	191	24.7	644	4	BG636414
15	191	24.7	654	4	BI374334
16	189	24.3	836	9	CNS0172X
17	188	24.3	642	4	BI173041
18	187	24.2	635	1	AI135629
19	184	23.8	600	4	BI163592
20	183	23.6	736	1	AA949890
21	182	23.5	694	4	BI354111
22	180	23.3	615	4	BI364613
23	179	23.1	655	4	BI484774
24	178	23.0	598	4	BI170396
25	176	22.7	585	1	AI133998
26	174	22.5	564	4	BI635692
27	170	22.0	658	4	BI374189
28	168	21.7	585	4	BI242003
29	167	21.6	539	4	BI373819
30	167	21.6	564	1	AI542327
31	166	21.4	649	4	BI162865
32	162	20.9	690	4	BG640860
33	160	20.7	554	4	BI167839
34	159	20.5	555	4	BI370013
35	155	20.0	646	4	BI170481
36	154	19.9	525	2	BF502679
37	148	19.1	539	1	AI107144
38	147	19.0	495	1	AI533215
39	147	19.0	527	4	BI586931
40	146	18.9	441	7	CO295609
41	141	18.2	424	7	CO337666
42	140	18.1	502	4	BI237763
43	138	17.8	414	7	CO295621
44	138	17.8	414	7	CO304185
45	138	17.8	414	7	CO312883

ALIGNMENTS

BF502026 810 bp mRNA linear EST 19-APR-2001
AT17637.5prime AT Drosophila melanogaster adult testes POFB7
Drosophila melanogaster cDNA clone AT17637 5 similar to OsetStt3:
FBan0007748 'enzyme' located on: 3R 96B16-96B17: 04/09/2001, mRNA
sequence.

ACCESSION BF502026 GI:13695997

VERSION EST.

KEYWORDS Drosophila melanogaster (fruit fly)

SOURCE Drosophila melanogaster

ORGANISM Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 810)

AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E.,

Berman,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V.,

Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N.,

Li,P., Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nuno,J.,

Pacileb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C.,

Lewis,S.E., Celniker,S. and Rubin,G.M.

BDGS/HMI AT Drosophila EST Project

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi:11585327.

Contact: Stapleton, M.

EDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu

hit genomic AE003569: arm.X [20544566-20835979]

estimated-cyto:19E4-19F1: 04/09/2001 hit P element 1(3)j299:

1(3)j2D9 AQ026308 inserted at base 292 5' end of P element Inverse
 PCR: 02/10/2001
 Plate: AT.176 row: D column: 1
 High quality sequence stop: 713.
 Location/Qualifiers

FEATURES

source

1. 810
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="AT11637"
 /sex="male"
 /dev_stage="0-3 day old Ore-R males"
 /lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
 AT.121-AT.319: DHS-alpha Tona"
 /clone_lib="AT Drosophila melanogaster adult testes pOTB7"
 /note="Organ: ADULT testes; Vector: pOTB7; Site:1: EcoRI;
 Site:2: XhoI; The mRNA for the testis library was made
 from testes and seminal vesicles hand dissected from 0-3
 day old Ore-R males. RNA kindly provided by the lab of
 Margaret Fuller. Sized fractionated cDNAs were directly
 ligated into pOTB7. Plasmid cDNA library."

ORIGIN

Alignment Scores:
 Pred. No.: 2,968-252 Length: 810
 Score: 245.00 Matches: 245
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.65% Indels: 0
 DB: 2 Gaps: 0

US-10-028-384-8 (1-774) x BF502026 (1-810)

Qy 1 MetAsnArgThrProTyrMetLeuAsnSerIysValAlaGlyTyrSerIleuThr 20
 Db 49 ATGAATCGGACCGCGGAGATCTCTGAACAGCAAGTGGCTGCTACAGCGCTATACACC 108
 Qy 21 PheAlaIleLeuLeuIleAlaTrrPheLeuAlaGlyPheSerArgLeuPheAlaValIle 40
 Db 109 TTCGGCATCTCTGTAATCGCTGCTGGCGGATTTTCCTCTGCTCTTCGCGCTATC 168
 Qy 41 ArgPheGluSerIleIleHisGluPheAspProTrrPheAsnTyrArgAlaThrAlaTyr 60
 Db 169 CGTTTCGATGCGATTTATCCATGAGTTTATCCGTTGTTCAACTACCGGCGCCACCGGCTAC 228
 Qy 61 MetValGlnAsnGlyTrrTrrAsnPheLeuAsnTrrPheAspGluArgAlaTrrTrrPro 80
 Db 229 ATGGTGCAGATGTTTGTACACTTCTCTCACTGTTTCACGAGCGGCGATGATCCG 288
 Qy 81 LeuGlyArgTrrIleValGlyGlyThrValTrrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 289 CTCGGCAGATGTTGGCGGTACCGTCTATCCCGCGCTGATGATTACGTCGGCGGAATC 348
 Qy 101 HisTrrPheLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 349 CATTCGCTGCTGCAGTACTCAACATACCGTTCATATTCGTGACATCTCGGTGTTCTTG 408
 Qy 121 AlaProlIlePheSerGlyLeuThrSerIleSerThrTrrIleuLeuThrIysGluLeuTrrP 140
 Db 409 GCGCGCATCTTCAGTGGCGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGGG 468
 Qy 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTrrIleSer 160
 Db 469 TCCGGCGGCGCGGCTCTTCGCGCGCATCTTCATCGCCATCGTGGCTGCTACATCACT 528
 Qy 161 ArgSerValAlaGlySerTrrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 Db 529 AGGTTCGGTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 588
 Qy 181 TrrPheLeuTrrValArgSerValIysThrGlySerValPheTrrPheAlaAlaAla 200
 Db 589 TACTTCCGTGGTGGCTGCTGATGAGACTGATGATGATGATGATGATGATGATGATGATG 648

Qy 201 LeuSerTrrPheTrrMetValSerAlaTrrGlyGlyTrrValPheIleAlaLeuIle 220
 Db 649 TTATCTTACTTCTTACATGTTGTCGCCCTGGGTGGTACGTTGTTCAATCACTGATA 708
 Qy 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTrrSerProArgLeuLeuThrSer 240
 Db 709 CCCTTCGACGCTTCTGTTACTGCTCATTATGGCAGTACTCGCGCGTGTCTGACCA 768
 Qy 241 TrrSerThrPheTrr 245
 Db 769 TACAGCACCTTCTAC 783

RESULT 2

BF503420 701 bp mRNA linear EST 19-APR-2001
 LOCUS AT19406 Sprime AT Drosophila melanogaster adult testes pOTB7
 DEFINITION Drosophila melanogaster cDNA clone AT19406 5 similar to OStStt3:
 Pfam0007748 'enzyme' located on: 3R 96B16-96B17; 04/09/2001, mRNA
 sequence.

ACCESSION

VERSION BF503420 GI:113697340

KEYWORDS

EST.

SOURCE

ORGANISM Drosophila melanogaster (fruit fly)

Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

REFERENCE

1 (bases 1 to 701)

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E.,
 Berlan, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V.,
 Farfan, D., Frise, B., George, R., Gonzalez, M., Guarin, H., Harris, N.,
 Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C. J., Nunoo, J.,
 Paclob, J., Paragosa, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,
 Lewis, S. E., Celniker, S. and Rubin, G. M.
 BGP/HMI AT Drosophila EST Project
 Unpublished (2000)

TITLE

JOURNAL

COMMENT

On Dec 6, 2000 this sequence version replaced gi:11596721.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit genomic AB003750: arm:3R [20671983,20899380]

estimated-cyto:96B10-96C1: 04/09/2001

Plate: AT.194 row: A column: 6

High quality sequence stop: 602.

FEATURES

source

1. 701

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="AT19406"

/sex="male"

/dev_stage="0-3 day old Ore-R males"

/lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
 AT.121-AT.319: DHS-alpha Tona"

/clone_lib="AT Drosophila melanogaster adult testes pOTB7"

/note="Organ: ADULT testes; Vector: pOTB7; Site:1: EcoRI;
 Site:2: XhoI; The mRNA for the testis library was made
 from testes and seminal vesicles hand dissected from 0-3
 day old Ore-R males. RNA kindly provided by the lab of
 Margaret Fuller. Sized fractionated cDNAs were directly
 ligated into pOTB7. Plasmid cDNA library."

ORIGIN

Alignment Scores:

Pred. No.: 1,758-226 Length: 701
 Score: 221.00 Matches: 221
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 28.55% Indels: 0
 DB: 2 Gaps: 0

```

US-10-028-384-8 (1-774) x BF503420 (1-701)

QY 224 ValPheValLeuLeuMetGlyArgTyrSerProArgLeuLeuThrsTyrSerThr 243
Db 2 GTCTTCGTACTCTCATTTATGGCAGGTAATCGCGCGTCTGTGACCACTACAGCACC 61
QY 244 PheTyrLeuLeuGlyLeuPheSerMetGlnIleProPheValGlyPheGlnProIle 263
Db 62 TTCTACATCTCGGACTCTGTTCTCCATGAGATCCCTTCGTTGGATTCACCAACGATA 121
QY 264 ArgThrSerGlnHisMetAlaLeuGlyValPheValLeuLeuMetAlaValAlaThr 283
Db 122 CGCACCACTGACATGCTGCGTGGAGTGTGTGCTCTTATGGCGCTGGCCACC 181
QY 284 LeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPheIleValGly 303
Db 182 TTGCGCCATTTGCATGCTGCTGCGCAACAGATTCGGAAGCTGTTCATCGTGGC 241
QY 304 GlyLeuValGlyValGlyValPheValAlaValValLeuThrsMetLeuGlyVal 323
Db 242 GGATTGCTGCTGGCGTGGCGTCTTTGTCGCCGCTCGTGTGCTCACCATGCTGGCGTT 301
QY 324 ValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHis 343
Db 302 GTGGCCCGTGGAGTGGAGCTTTCTACCTGCTGTGGATCTGCTACGCAAGATCCAC 361
QY 344 IleProIleAlaSerValSerGlnHisGlnProThrThrTrpPheSerPhePhe 363
Db 362 ATTCCCATCAATGCAATCGTGTGGAGCATCAGCCACCACCTTGTTCTGTTCTCTTT 421
QY 364 AspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIleLysGlnIle 383
Db 422 GATCTGCACATCTCTGTGTGGCTTCCAGTGGAGTGGTACTGCATCAAGCAATC 481
QY 384 AsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyVal 403
Db 482 AACGACGACGCGCTTTCTGTGTGCTGTGACCCATCAGTGGCGTTTACTTCGTGTGTG 541
QY 404 MetValArgLeuMetLeuThrsLeuThrsProValValCysMetLeuAlaGlyValAlaPhe 423
Db 542 ATGGTGGCTTGTATGTTACCTCAGCCGCTGTGTGTCATGCTGGCGAGTGGCGCTTT 601
QY 424 SerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIle 443
Db 602 TCGGACATGTTGATGTCTCTGCAAGAGGATTCGTTAAGCAATGGGCACAGCCATA 661
QY 444 Ser 444
Db 662 AGC 664

RESULT 3
BF496296
LOCUS
DEFINITION
US-10-028-384-8 (1-774) x BF496296 (1-655)
EST
AT10060.5 Prime AT Drosophila melanogaster adult testes pOTB7
Drosophila melanogaster cDNA clone AT10060.5 similar to OstStt3:
FBan007748 'enzyme' located on: 3R 96B16-96B17; 04/08/2001, mRNA
sequence.
BF496296
BF496296.2 GI:13693786
EST
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila
1 (bases 1 to 655)
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E.,
Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V.,
Farfan, D., Frise, E., George, R., Gonzalez, M., Guatin, H., Harris, N.,
Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J.,
Pacib, J., Paragosa, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,
Lewis, S.E., Ceiniker, S. and Rubin, G.M.

TITLE
JOURNAL
COMMENT
BDGP/HMI AT Drosophila EST Project
Unpublished (2000)
On Dec 6, 2000 this sequence version replaced gi:11579597.
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AEO03750: arm:3R [20671983,20899380]
estimated-cyto:96B10-96C1: 04/08/2001 hit P element 1(3)j2D9:
1(3)j2D9 AQ026308 inserted at base 292 5', end of P element Inverse
PCR: 02/09/2001
Plate: AT 100 row: E column: 12
High quality sequence stop: 577.
Location/Qualifiers
1..655
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="AT10060"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT 10-AT 120: DH5-alpha. Plates
AT 121-AT 319: DH5-alpha 10mA"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/note="Organ: ADULT testes; Vector: pOTB7; Site: 1: EcoRI;
Site 2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."

ORIGIN
Alignment Scores:
Pred. No.: 5,59e-219 Length: 655
Score: 214.00 Matches: 214
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.65% Indels: 0
Db: 2 Gaps: 0

US-10-028-384-8 (1-774) x BF496296 (1-655)
QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
Db 14 ATGATCGGACCGCGAAGATGCTGACACAGAGTGGCTGGCTACAGAGCCTAATCACC 73
QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db 74 TTCGCCATCTCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTTCGCCGTCATC 133
QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
Db 134 CGTTTCGAGTCGATATCCATCATGAGTTGATCCGCTGGTTCACTACCGGCCACCGCTAC 193
QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
Db 194 ATGGTGCAGATGGTGTGTTGTTCAACTCTCTCACTGGTTCGACGAGCGCATGTATCCG 253
QY 81 LeuGlyArgIleValGlyValThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db 254 CTGGCAGGATTTGGCGGCTACCGTCTATCCCGGCTGATGATTACGTCGGCGGATC 313
QY 101 HistTrpLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db 314 CATGGCTGTGTCAGCTACTCAACATACCGGTCCATATTGTCGACATCTCGCTGTTCTCG 373
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
Db 374 CGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACTCTGACCAAGAGGCTGTGG 433
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160

```

```

Db      434 TCCGGCGCGCGCGCTTTCGCGCGCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT 493
Qy      161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db      494 AGGTCGGTGGCTGGATCGTACGATAACGAGGGCAATGCCATATTCGCCCTGCAGTTCCAC 553
Qy      181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheThrSerAlaAlaAla 200
Db      554 TACTTCCTGGTGGGTCGCTCAGTGAGACATGGATCCGTTTCGGTGGCGCCAGCCGCT 613
Qy      201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrVal 214
Db      614 TTATCTACTTCTACATGATGGTCCGCTGGGCTGGCTACGCTG 655

RESULT 4
LOCUS   CK657451 715 bp mRNA linear EST 30-JAN-2004
DEFINITION Drosophila melanogaster larval-early pupal pOT2
ACCESSION CK657451
VERSION   CK657451.1 GI:41400976
KEYWORDS EST.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 715)
AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
          Lewis,S. and Rubin,G.M.
TITLE     BDGP/HMI Drosophila EST Project
JOURNAL   Unpublished (2001)
COMMENT   Contact: Stapleton, M.
          BDGP
          Lawrence Berkeley National Lab
          One Cyclotron Rd, Berkeley, CA 94720, USA
          Fax: 510 486 6798
          Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
          Plate: LP.237 row: E column: 11
          High quality sequence stop: 644.
          Location/Qualifiers
            1..715
              /organism="Drosophila melanogaster"
              /mol_type="mRNA"
              /db_xref="taxon:7227"
              /clone="LP23759"
              /sex="male and female"
              /dev_stage="larvae-pupae"
              /lab_host="DHS-alpha"
              /clone_lib="LP Drosophila melanogaster larval-early pupal
              pOT2"
              /note="Organ: whole body; Vector: pOT2; Site 1: EcORI;
              Site 2: XhoI; Sized fractionated cDNAs were directly
              ligated into pOT2. Plasmid cDNA library."

```

FEATURES

```

source
1..715
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="LP23759"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DHS-alpha"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pOT2"
/note="Organ: whole body; Vector: pOT2; Site 1: EcORI;
Site 2: XhoI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."

ORIGIN
Alignment Scores:
Pred. No.:      6,18-219      Length:      715
Score:          214.00      Matches:      214
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      27.65%      Indels:      0
Db:              7          Gaps:      0

US-10-028-384-8 (1-774) x CK657451 (1-715)

```

```

Qy      1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeulleThr 20
Db      60 ATGAATCGGACCGCCGAGATGCTGAACAGCAAGGTGGCTGGCTACAGCAAGCTTAATCACC 119
Qy      21 PheAlaIleLeuLeulleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaIle 40

```

```

Db      120 TTCCGCATCTCTCTAATCGCTGGCTGGCGGAGATTTCTCTCCCTCTTTCGGCGTCATC 179
Qy      41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
Db      180 CGTTTCGAGTCGATATTCATCGTGGTTTCAACTACCGGGCCACCGGCTAC 239
Qy      61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
Db      240 ARGGTGCAGATGGTGGTACAACTTCTCACTGGTTCGACGAGCCGCGATGGTATCCG 299
Qy      81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db      300 CTCCGCAGGATTTGGGGCGGTACCGTCTATCCGCGCTGATGATTACGTCGCGGGAATC 359
Qy      101 HisTrpLeuLeuHisValIleuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db      360 CATTGGCTCTCCAGTACTCAACATACCGGTCCATATTCTGTGACATCTGGGTTCCTG 419
Qy      121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrIleuLeuThrLysGluLeuTrp 140
Db      420 GCGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 479
Qy      141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db      480 TCCGCGGGCGCGGCTCTTTCGCGCCAGCTTCATCCGCATCTGCTGCTGCTACATCAGT 539
Qy      161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db      540 AGGTGGTGGCTGGATCGTACGATAACGAGGCGCATTCGCATATTTCGCCCTGCAGTTCCACC 599
Qy      181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheThrSerAlaAlaAla 200
Db      600 TACTTCTCTGGTGGCTGGCTCAGTGAAGACTGATCGGTGTTCTGGTGGCGCGGAGCCGCT 659
Qy      201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrVal 214
Db      660 TTGTCTCTTCTACATGGTGTCCGCTGGGCTGGCTACGCTG 701

RESULT 5
LOCUS   CK657649 679 bp mRNA linear EST 30-JAN-2004
DEFINITION Drosophila melanogaster larval-early pupal pOT2
ACCESSION CK657649
VERSION   CK657649.1 GI:41401174
KEYWORDS EST.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 679)
AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
          Lewis,S. and Rubin,G.M.
TITLE     BDGP/HMI Drosophila EST Project
JOURNAL   Unpublished (2001)
COMMENT   Contact: Stapleton, M.
          BDGP
          Lawrence Berkeley National Lab
          One Cyclotron Rd, Berkeley, CA 94720, USA
          Fax: 510 486 6798
          Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
          Plate: LP.241 row: B column: 7
          High quality sequence stop: 670.
          Location/Qualifiers
            1..679
              /organism="Drosophila melanogaster"
              /mol_type="mRNA"
              /db_xref="taxon:7227"
              /clone="LP24119"
              /sex="male and female"
              /dev_stage="larvae-pupae"

```

/lab_host="DHS-alpha"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pOT2"
/note="Organ: whole body; Vector: pOT2; Site 1: EcORI;
Site 2: XhoI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."

ORIGIN

Alignment Scores:

Pred. No.: 1.66e-212 Length: 679
Score: 208.00 Matches: 208
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.87% Indels: 0
DB: 7 Gaps: 0

US-10-028-384-8 (1-774) x CK657649 (1-679)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
DB 54 ATGAATCGAGCGCGAGATGCTGAACAGCAAGGTGGCTGCTACAGCAGCTAATCACC 113
QY 21 PheAlaIleLeuLeuIleAlaTyrPheSerSerArgLeuPheAlaValIle 40
DB 114 TTGCGCATCTGCTAATCGCTGCTGCGGATTTTCTCTGCTCTTTCGCGGTATC 173
QY 41 ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60
DB 174 CGTTTCAGTTCGATATTCATGAGTTTGTATCCGTGGTTCAACTACCGGCGCACCGCTAC 233
QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrTyrPro 80
DB 234 ATGCTGCAGATGCTTGGTACACTTCTCACTGGTTCAGCAGCGCATGTTATCCG 293
QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyIle 100
DB 294 CTGCGCAGGATTTGGGCGGTACCGTCTATCCGGCGCTGATGATTACGTCGCGCGAATC 353
QY 101 HisTyrLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
DB 354 CATTTGGTGTGTCAGTACTCAATACCGGTCCATTCGTGACATCTGCTGCTGCTGCTG 413
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140
DB 414 GCGCGCATCTTCACTGCGCTGACTCCATCTCCACCTACCTGCTGACCAAGAGCTGTGG 473
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
DB 474 TCCGCGGCGCGCGCTTTCGCGCGCAGCTTCATCGCATCTGCTGCTGCTGCTGCTG 533
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
DB 534 AGTGTGCTGCTGATCGTACGATACAGAGGATTTGCCATATTCGCTGCTGCTGCTGCT 593
QY 181 TyrPheLeuTyrValArgSerValLysThrGlySerValPheTyrSerAlaAlaAla 200
DB 594 TACTTCTGTGGGTGGCTCAGTGAAGACTGGATCGGTCTTCTGCTGCTGCTGCTGCTG 653
QY 201 LeuSerTyrPheTyrMetValSer 208
DB 654 TTGCTCTACTTCTACATGTTGTTCC 677

RESULT 6

CK659033 681 bp mRNA linear EST 30-JAN-2004
LOCUS LP16462.5prime LP Drosophila melanogaster larval-early pupal pOT2
DEFINITION Drosophila melanogaster cDNA clone LP16462.5, mRNA sequence.
ACCESSION CK659033
VERSION CK659033.1 GI:41402558
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Phylophoridae; Drosophilidae; Drosophila.
1 (bases 1 to 681)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
Plate: LP164 row: F column: 2

High quality sequence stop: 670.
Location/Qualifiers
1. 681

FEATURES

/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="LP16462"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DHS-alpha"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pOT2"
/note="Organ: whole body; Vector: pOT2; Site 1: EcORI;
Site 2: XhoI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."

ORIGIN

Alignment Scores:
Pred. No.: 1.67e-212 Length: 681
Score: 208.00 Matches: 208
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.87% Indels: 0
DB: 7 Gaps: 0

US-10-028-384-8 (1-774) x CK659033 (1-681)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
DB 56 ATGAATCGAGCGCGAGATGCTGAACAGCAAGGTGGCTGCTACAGCAGCTAATCACC 115
QY 21 PheAlaIleLeuLeuIleAlaTyrPheSerSerArgLeuPheAlaValIle 40
DB 116 TTGCGCATCTTCTAATCGCTGCTGCGGATTTTCTCTGCTCTTTCGCGGTATC 175
QY 41 ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60
DB 176 CGTTTCAGTTCGATATTCATGATCCGTGGTTCAACTACCGGCGCACCGCTAC 235
QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrTyrPro 80
DB 236 ATGTTGCAGATGTTGGTACCACTTCTCACTGGTTCAGCAGCGCGCATGATATCCG 295
QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyIle 100
DB 296 CTGCGCAGGATTTGGGCGGTACCGTCTATCCGGCGCTGATGATTACGTCGCGCGAATC 355
QY 101 HisTyrLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
DB 356 CATTTGGTGTGTCAGGACTCAACATCCGCTCATATTCGTGACATCTCGGTGTTCTG 415
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140
DB 416 GCGCGCATCTTCACTGCGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGTGG 475
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160

```

Db      476 TCCGCGCGCGCGCTCTTCGCGCGCAGCTTCATCGCCATCGTGGCTGCTACATCAGT 535
Qy      161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db      536 AGTTCGGTGGCTGGATCTGATGATACACGAGGCGATTGCCATATTCGCTCGAGTTCACC 595
Qy      181 TyrPheLeuTyrValArgSerValGlyThrGlySerValPheThrSerAlaAlaAla 200
Db      596 TACTTCTGTGGTGGCTGCTCAGTGAAGACTGGATCCGTGTTCTGCTGGCGCGCAGCGCT 655
Qy      201 LeuSerTyrPheTyrMetValSer 208
Db      656 TTGTCTACTTCTACATGTTGCC 679

RESULT 7
Bi171940
LOCUS
DEFINITION
  Bi171940 652 bp mRNA linear EST 09-JUL-2001
  RE13841.5prime RE Drosophila melanogaster normalized Embryo pflc-1
  Drosophila melanogaster cDNA clone RE13841.5 similar to OstStt3:
  FBan0007748 'enzyme' located on: 3R 96B16-96B17; 04/06/2001, mRNA
  sequence.
ACCESSION
  Bi171940
KEYWORDS
  Bi171940.1 GI:14637747
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 652)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champagne,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacieb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST\_est@fruitfly.berkeley.edu
  hit genomic AE003750: arm:3R [20671983,20899380]
  estimated-cyto:96B10-96C1: 04/06/2001 hit P element 1(3)J2D9:
  1(3)J2D9 AQ26308 inserted at base 292 5' end of P element Inverse
  PCR: 04/06/2001
  Plate: RE.138
  High quality sequence stop: 590.
  Location/Qualifiers
    FEATURES
      source
        1..652
        /organism="Drosophila melanogaster"
        /mol_type="mRNA"
        /db_xref="taxon:7227"
        /clone="RE13841"
        /sex="male and female"
        /dev_stage="0-24 hours mixed stage embryonic"
        /lab_host="DHS-alpha TorA"
        /clone_lib="RE Drosophila melanogaster normalized Embryo
        pflc-1"
        /note="Organ: embryo; Vector: pflc1; Site:1: XhoI; Site 2:
        BamHI; Library was kindly generated by Piero Carninci at
        the RIKEN. The library was normalized and excised using
        Cre recombinase. Plasmid cDNA library."
      ORIGIN
        Alignment Scores:
        Pred. No.: 1,11e-200 Length: 652
        Score: 197.00 Matches: 197
        Percent Similarity: 100.00% Conservative: 0
        Best Local Similarity: 100.00% Mismatches: 0
        Query Match: 25.45% Indels: 0

```

```

Db      4 Gaps: 0
US-10-028-384-8 (1-774) x Bi171940 (1-652)
Qy      1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
Db      60 ATGAATCGGAGCGCCGAGATGCTGAACAGCAAGTGGCTGGCTACAGCAGCTAATCACC 119
Qy      21 PheAlaIleLeuLeuIleAlaTyrPheAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db      120 TTGCGCATCTCTAATCGCTGGCTGGCCGGAATTTCTCTCGCCCTTCGCGGTGATC 179
Qy      41 ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60
Db      180 CGTTTCGAGTCGATTATCAATGAGTTTCATCGGTGGTTCAACTACCGGCGCCACCGCTAC 239
Qy      61 MetValGlnAsnGlyTyrTyrPheLeuAsnTyrPheAspGluAlaTyrTyrPro 80
Db      240 ATGGTGCAGAAATGGTTGTGTACAACTTCCTCAACTGGTTTCGACGAGCGCGCATGGTATCCG 299
Qy      81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db      300 CTCGGCAGGATTTGTGGCGGTACCGTCTATCCGGCCCTGATGATTACGTCGCGGGAATC 359
Qy      101 HistTyrLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db      360 CATTCGCTGCTGCACCTACTCAACATACCGGTCCATATTCGTGCATCTGGGTGTCCTG 419
Qy      121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140
Db      420 GCGCCGATCTTCAGTGGCTGACCTCCATCTCCACTTCCACTGCTGTCACGAGAGCTGG 479
Qy      141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db      480 TCCGCGGCGCGCGCTCTTCGCGCGCAGCTTCATCGCCATCGTGCCTGGCTACATCAGT 539
Qy      161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db      540 AGTTCGGTGGCTGGATCGTACGATAACGAGGCGGATTCGCAATTCGCCCTGAGTTCACC 599
Qy      181 TyrPheLeuTyrValArgSerValTyrThrGlySerValPheThrSerAla 197
Db      600 TACTTCTGTGGTGGCTGCGCTCAGTGAAGACTGGATCGTGTCTGTGTCGCGCC 650

Bi357074 666 bp mRNA linear EST 31-JUL-2001
RE43425.5prime RE Drosophila melanogaster normalized Embryo pflc-1
Drosophila melanogaster cDNA clone RE43425.5 similar to OstStt3:
FBan0007748 'enzyme' located on: 3R 96B16-96B17; 05/13/2001, mRNA
sequence.
ACCESSION
  Bi357074
VERSION
  Bi357074.1 GI:15051528
KEYWORDS
  EST.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 666)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champagne,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacieb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798

```


Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit genomic AB003750: arm:3R [20671983,20899380]
 estimated-cyto:96B10-96C1: 05/13/2001 hit P element l(3)j2D9:
 1(3)j2D9 A0026308 inserted at base 292 5' end of P element inverse
 PCR: 05/13/2001
 Plate: RE:434 row: C column: 1
 High quality sequence stop: 563.
 Location/Qualifiers

FEATURES

1..666
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE43425"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Alignment Scores:
 Pred. No.: 1,358-199 Length: 666
 Score: 196.00 Matches: 196
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 25.32% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x B1357074 (1-666)

QY	1	MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr	20
Db	79	ATGAATCGGACCGCGAAGATCTGAACAGCAAGTGGCTGCTACAGACGCTTAATCACC	138
QY	21	PheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValle	40
Db	139	TTCCGCATCTGCTTAATCGCGCTGGCTGGCGGATTTTCCTCGCCCTCTTCGCGCTCATC	198
QY	41	ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr	60
Db	199	CGTTTCGAGTGATTAATCCATGAGTTTGATCCGGTGGTTCACTACCGGCGCACCGCTAC	258
QY	61	MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro	80
Db	259	ATGGTGAGAAATGGTTGGTACAACTCTCAACTGGTTTCGACGAGCGCGCATGGTATCCG	318
QY	81	LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle	100
Db	319	CTCGGCAAGATTGGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGCGAATC	378
QY	101	HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120
Db	379	CATTGGCTGTGCGAGTACTCAACATACCGGTCCATATTCGTGACATCTGGGTTCCTG	438
QY	121	AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp	140
Db	439	CGCGCGATCTTCAGTGGCGCTGACCTCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG	498
QY	141	SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer	160
Db	499	TCGCGGCGCGCGCTTCGCGCGAGTTTCATGCGCATCTGCTGGCTACATCAT	558
QY	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180
Db	559	AGTCGCGTGGCTGGATCGTACGATAACAGAGGCAATTCGCATATTCGCCCTGCAGTTTACC	618
QY	181	TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer	196
Db	619	TACTTCTGTGGGTGGCTGCTGAGAGACTGGATCGGTGTTCGGTCG	666

RESULT 9

B1227902
 LOCUS

DEFINITION

B1227902 668 bp mRNA linear EST 11-JUL-2001
 RE25288.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
 Drosophila melanogaster cDNA clone RE25288 5 similar to OatStt3:
 FBan0007748 'enzyme' located on: 3R 96B16-96B17: 04/12/2001, mRNA
 sequence.

ACCESSION

VERSION B1227902.1 GI:14695166

KEYWORDS

SOURCE EST.

ORGANISM

Drosophila melanogaster (fruit fly)
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 668)

AUTHORS

Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
 Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
 George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
 Misra,S., Mungall,C.J., Nuno,J., Pacleb,J., Paragas,V., Park,S.,
 Phananavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
 Rubin,G.M.

TITLE

BDGP/HMI RE Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit genomic AB003750: arm:3R [20671983,20899380]

estimated-cyto:96B10-96C1: 04/12/2001 hit P element l(3)j2D9:

1(3)j2D9 A0026308 inserted at base 292 5' end of P element inverse

PCR: 03/20/2001 row: H column: 4

Plate: RE:252 row: H column: 4

High quality sequence stop: 520.

FEATURES

-source

1..668

Location/Qualifiers

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="RE25288"

/sex="male and female"

/dev_stage="0-24 hours mixed stage embryonic"

/lab_host="DHS-alpha Tona"

/clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"

/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Alignment Scores:

Pred. No.: 1,358-199 Length: 668
 Score: 196.00 Matches: 196
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 25.32% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x B1227902 (1-668)

QY	1	MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr	20
Db	81	ATGAATCGGACCGCGAAGATCTGAACAGCAAGTGGCTGCTACAGACGCTTAATCACC	140
QY	21	PheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValle	40
Db	141	TTCCGCATCTGCTTAATCGCGCTGGCTGGCGGATTTTCCTCGCCCTCTTCGCGCTCATC	200
QY	41	ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr	60

Db	201	CGTTTCGAGTCGATTATCATGAGTTTTCATCCGTGGTTCAACTACCGGGCCACCGCCCTAC	260
Qy	61	MetValGlnAsnGlyTyrTrpAsnPheLeuAsnTrpPheAspGluArgAlaTyrTrpPro	80
Db	261	ATGGTCGAGAAAGTGGTGGTACAACTTCCTCACTGGTTTCGACGAGCGGCGATGGTATCCG	320
Qy	81	LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyLeu	100
Db	321	CTCGGCGAGGATTGTGGCGGTACCGTCTATCCCGCCCTCATGATTCGCGGGGGAATC	380
Qy	101	HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120
Db	381	CATTGGCTGCTGCACGTACTCAACATACCGTTCATATTCGTGACATCTGGTGTCTCTG	440
Qy	121	AlaProilePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIleGluLeuTrp	140
Db	441	GGCCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG	500
Qy	141	SerAlaGlyAlaGlyLeuPheAlaAsnSerPheIleAlaIleValProGlyTyrIleSer	160
Db	501	TCGCGGGCGCGGCGCTTCGCGCGGACGCTTCATCGCCATCTGCTGCTGCTGCTGCTG	560
Qy	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180
Db	561	AGTTCGGTGGTGGATCGTACGATACGAGGCGATTCGATATTCGCGCTGCGATTCACC	620
Qy	181	TyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSer	196
Db	621	TACTTCCTGGTGGTGGCTCAGTGAAGACTGGATCGTGGTTCGTGGTGG	668
RESULT 10			
LOCUS	AI257750	641 bp mRNA linear EST 19-APR-2001	
DEFINITION	LP06212.5prime LP Drosophila melanogaster larval-early pupal POT2		
ACCESSION	Drosophila melanogaster cDNA clone LP06212 5prime, mRNA sequence.		
VERSION	AI257750		
KEYWORDS	AI257750.1 GI:3865275		
SOURCE	EST.		
ORGANISM	Drosophila melanogaster (fruit fly)		
REFERENCE	Drosophila melanogaster		
AUTHORS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
TITLE	1 (bases 1 to 641)		
JOURNAL	Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S. and Rubin, G.M.		
COMMENT	BDGP/HMI Drosophila EST Project Unpublished (2001) Contact: Stapleton, M.		
FEATURES			
Source			
1..641			
/organism="Drosophila melanogaster"			
/mol_type="mRNA"			
/db_xref="taxon:7227"			
/clone="LP06212"			
/sex="male and female"			
/dev_stage="larvae-pupae"			
/lab_host="DH5-alpha"			
/clone_lib="LP Drosophila melanogaster larval-early pupal POT2"			
/note="Organ: whole body; Vector: POT2; Site: 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."			
ORIGIN			
Alignment Scores:			
Pred. No.:	1,558-198	Length:	641
Score:	195.00	Matches:	195
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	25.19%	Indels:	0
DB:	1	Gaps:	0
US-10-028-384-8 (1-774) x AI257750 (1-641)			
Qy	1	MetAsnArgThrProIleAsnSerIleValAlaGlyTyrSerSerIleIleThr	20
Db	55	ATGAAATCGGACCGCGAAGATGCTGAACAGCAAGGTGGCTACAGAGCGCTAATCACC	114
Qy	21	PheAlaIleIleLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle	40
Db	115	TTCCGCATCTCTGCTAATCGCCTGGCTGGCCGGAATTTCTCTCGCTCTTCGCCCTCATC	174
Qy	41	ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr	60
Db	175	CGTTTCGAGTCGATTATCCATGAGTTTGATCCGTGGTTCAACTACCGGGCCACCGCCTAC	234
Qy	61	MetValGlnAsnGlyTyrTrpAsnPheLeuAsnTrpPheAspGluArgAlaTrpTrpPro	80
Db	235	ATGGTCGAGAAATGGTGGTACAACTTCCTCACTGGTTCGACGCGCGCATGGTATCCG	294
Qy	81	LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyLeu	100
Db	295	CTGGGAGGATTTGGGCGGTACCGTCTATCCGGCTGATGATTACGTCGCGGGAATC	354
Qy	101	HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120
Db	355	CATTGGCTGTCGACGTACTCAACATACCGGTCCATATTCGTGACATCTGCGGTTCCTG	414
Qy	121	AlaProilePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIleGluLeuTrp	140
Db	415	GGCCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG	474
Qy	141	SerAlaGlyAlaGlyLeuPheAlaAsnSerPheIleAlaIleValProGlyTyrIleSer	160
Db	475	TCGCGGGCGCGGCGCTTCGCGCGGACGCTTCATCGCCATCTGCTGCTGCTGCTGCTG	534
Qy	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180
Db	535	AGTTCGGTGGTGGATCGTACGATACGAGGCGCATTCGCCATATTCGCCCTGCGATTCACC	594
Qy	181	TyrPheLeuTrpValArgSerValIleThrGlySerValPheTrp	195
Db	595	TACTTCCTGGTGGTGGCTCAGTGAAGACTGGATCGTGGTTCGTGG	639
RESULT 11			
LOCUS	BG641172	663 bp mRNA linear EST 23-APR-2001	
DEFINITION	SD12448.5prime SD Drosophila melanogaster Schneider L2 cell culture		
ACCESSION	POT2 Drosophila melanogaster cDNA clone SD12448 5 similar to		
VERSION	04/13/2001, mRNA sequence.		
KEYWORDS	EST.		
SOURCE	Drosophila melanogaster (fruit fly)		
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	1 (bases 1 to 663)		
TITLE	Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S. and Rubin, G.M.		
JOURNAL	BDGP/HMI Drosophila EST Project Unpublished (2001) Contact: Stapleton, M.		
COMMENT			

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

42	PheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyrMet	61
153	TTGAGTCGATATCACTAGTTGATCCGTTGGTTCACTACCGGCGCACCGCTCATATG	212
62	ValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrTyrProLeu	81
213	GTGCAGAAATGGTGTGTACAACTTCCTCAACTGGTTCGACGAGCGCGATGATATCCGCTC	272
82	GlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIleHis	101
273	GCCAGATTGTGGGCGGTACCGTCTATCCCGCCTGATGATAGTCCGGGGATCCAT	332
102	TyrLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeuAla	121
333	TGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATCTGCTGCTTCTCGCG	392
122	ProIlePheSerGlyLeuThrSerIleSerThrTyrTyrLeuLeuThrLysGluLeuTyrSer	141
393	CCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGTCC	452
142	AlaGlyAlaGlyLeuPheAlaIleSerPheIleAlaIleValProGlyTyrIleSerArg	161
453	GCGGGCGCGCGCTCTTCGCGCGCAGCTTCATCGGCATCGTGCCTGGCTACATCAGTAGG	512
162	SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr	181
513	TCGGTGGCTGGATCGTACATACAGGCGCATTCGCATATTCGCCCTCGATTCACCTAC	572
182	PheLeuTyrValArgSerValIlyThrGlySerValPheTyrP	195
573	TTCTGTGGTGGCTCAGTGAGACTGGATCCGTTCTGG	614

ULT 13

AI295381

LOCUS

DEFINITION

LP08987.5prime LP Drosophila melanogaster larval-early pupal pOT2

Drosophila melanogaster cDNA clone LP08987 5prime, mRNA sequence.

AI295381

AI295381.1

GI:3944788

EST.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 630)

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S. and Rubin, G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit P element sequence 1(3)2D9

Plate: 89 row: H column: 3

High quality sequence stop: 498.

Location/Qualifiers

1. .630

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP08987"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DHS-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal

pOT2"

FEATURES

source

1. .630

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP08987"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DHS-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal

pOT2"

REFERENCE

AUTHORS

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S. and Rubin, G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit P element sequence 1(3)2D9

Plate: 89 row: H column: 3

High quality sequence stop: 498.

Location/Qualifiers

1. .630

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP08987"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DHS-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal

pOT2"

Alignment Scores:	3.07e-194	Length:	630
Pred. No.:	191.00	Matches:	191
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	24.68%	Indels:	0
DB:	1	Gaps:	0
US-10-028-384-8 (1-774) x AI295381 (1-630)			
Qy	1	MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr	20
Db	56	ATGAATCGGACGCCGGAAGATGCTGAACGCAAGTGGTGGCTACAGCAGCTTAATCACC	115
Qy	21	PheAlaIleLeuLeuIleAlaTyrPheAlaGlyPheSerSerArgLeuPheAlaValIle	40
Db	116	TTGCCATCTCTGTAATACGCTGGCTGGCGGATTTTCCTCTCGCCTTTCGGCGGTCATC	175
Qy	41	ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr	60
Db	176	CGTTTCGAGTCGATATTCATGAGTTTGATCCGTTGTTCACTACCGGGCCACCGGCTAC	235
Qy	61	MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrTyrPro	80
Db	236	ATGGTGCAGAAATGGTGTGTAACAATCTCTCAACTGGTTCGACGCGCGCATGGTATCCG	295
Qy	81	LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle	100
Db	296	CTCGGCAGGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTAGTCCGGCGGAATC	355
Qy	101	HisTyrPheLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120
Db	356	CATTGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCCGTG	415
Qy	121	AlaProIlePheSerGlyLeuThrSerIleSerThrTyrTyrLeuLeuThrLysGluLeuTyr	140
Db	416	GCGCGGATCTTCAGTGGCGCTGACCTCCATCTCCACCTACTCTGTGACCAAGGAGCTGG	475
Qy	141	SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer	160
Db	476	TCCGCGGCGCGCGGCTCTTCGCGCGCAGCTTCATCGCCATCTCGTCCGTGCTACATCAGT	535
Qy	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180
Db	536	AGTTCGGTGGCTGGATCGTACGATACAGAGGCATTCGCATATTCGCCCTGCATTCACC	595
Qy	181	TyrPheLeuTyrValArgSerValIlyThrGly	191
Db	596	TACTTCTCTGTGGTGGCTCAGTCAAGACTGGA	628

RESULT 14

BG636414

LOCUS

DEFINITION

SD14123.5prime SD Drosophila melanogaster Schneider L2 cell culture

pOT2 Drosophila melanogaster cDNA clone SD14123 5 similar to

OstStt3: FBan007748 'enzyme', located on: 3R 96B16-96B17;:

04/13/2001, mRNA sequence.

BG636414

BG636414.1

GI:13763951

EST.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 644)

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S., and Rubin, G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit P element sequence 1(3)2D9

Plate: 89 row: H column: 3

High quality sequence stop: 498.

Location/Qualifiers

1. .630

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP08987"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DHS-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal

pOT2"

REFERENCE

AUTHORS

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S., and Rubin, G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit P element sequence 1(3)2D9

Plate: 89 row: H column: 3

High quality sequence stop: 498.

Location/Qualifiers

1. .630

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP08987"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DHS-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal

pOT2"

Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
One Cyclotron Rd., Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003750: arm:3R [20671983,20899380]
estimated-cyto:96B10-96C1: 04/13/2001 hit P element 1(3)j2D9:
1(3)j2D9 AQ026308 inserted at base 292 5', end of P element Inverse
PCR: 03/20/2001
Plate: SD.141 row: B column: 11
High quality sequence stop: 612.
Location/Qualifiers

FEATURES
source
1..644
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="SD14123"
/lab_host="DH5-alpha"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pOR2"
/note="Vector: pOR2; Site 1: EcoRI; Site 2: XhoI; Sized
fractionated cDNAs were directly ligated into pOR2.
Plasmid cDNA library."

ORIGIN
Alignment Scores:
Pred. No.: 3,149-194 Length: 644
Score: 191.00 Matches: 191
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.68% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x BG636414 (1-644)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20
DB 71 ATGAATCGAGCCGACGAGTGTGAAACAGCAAGGTGGCTGCTACAGCAGCTTAATCACC 130
QY 21 PheAlaLeuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValile 40
DB 131 TTCGCATCCTCTTAATCGCTGGCTGGCCGATTTTCCTCCCTCTTCGCGCTATC 190
QY 41 ArgPheGluSerIleIleGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
DB 191 CGTTTCGAGTCGATTATCATGAGTTTCATCCGTTGTTCAACTACCGGGCCACGCGCTAC 250
QY 61 MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
DB 251 ATGGTGCAATGGTGTGGTACAACTTCCTCAACTGGTTCGACGAGCGCGATGATCCG 310
QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyile 100
DB 311 CTCCGAGGATTGTGGGGTACCTCTATCCGCGCTGATGATTACGTCGCGGGAATC 370
QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
DB 371 CATTGGCTCTCGAGCTACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCTTG 430
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
DB 431 GCGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 490
QY 141 SerAlaGlyAlaGlyLeuPheAlaIleAlaSerPheIleAlaIleValProGlyTyrIleSer 160
DB 491 TCAGCGGGCGCGGCTCTTCGCGCCAGCTTCATCGCCATCGGTGGCTGCATCATCAGT 550
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180

551 AGTCGGTGGCTGATCGTACGATACGAGGCGGATTCGCTATTCGCTCGAGTTCCACC 610
181 TyrPheLeuTrpValArgSerValLysThrGly 191
611 TACTTCTGTGGTGGCTCGCTCAGTGAAGACTGGA 643

RESULT 15
B374334
LOCUS
DEFINITION
B1374334 654 bp mRNA linear EST 01-AUG-2001
REG1893.3Prime RE Drosophila melanogaster normalized Embryo pF1c-1
Drosophila melanogaster cDNA clone REG1893 5 similar to OstStt3:
FBAN007748 GO: [colligasecarboxylase transferase (GO:0004576); enzyme
(GO:0003824)] located on: 3R 96B16-96B17; 05/16/2001, mRNA
sequence.
ACCESSION
B1374334
VERSION
B1374334.1 GI:15070362
KEYWORDS
EST.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 654)
AUTHORS
Stapleton, M., Brockstein, P., Hong, L., Tyler, D., Berman, B.,
Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
Misra, S., Mungall, C. J., Nunoo, J., Pacieb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
Rubin, G.M.
TITLE
BDGP/HMI RE Drosophila EST Project
JOURNAL
Unpublished (2001)
COMMENT
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd., Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003750: arm:3R [20671983,20899380]
estimated-cyto:96B10-96C1: 05/16/2001 hit P element 1(3)j2D9:
1(3)j2D9 AQ026308 inserted at base 292 5', end of P element Inverse
PCR: 05/16/2001
Plate: RE.618 row: H column: 9
High quality sequence stop: 595.
Location/Qualifiers

FEATURES
source
1..654
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE61893"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pF1c-1"
/note="Organ: embryo; Vector: pF1c1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN
Alignment Scores:
Pred. No.: 3,199-194 Length: 654
Score: 191.00 Matches: 191
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.68% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x B1374334 (1-654)

QY 1 MethAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20
DB 81 ATGAATCGAGCCGACGAGTGTGAAACAGCAAGGTGGCTGCTACAGCAGCTTAATCACC 140

Thu Dec 16 16:25:15 2004

QY	21	PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle	40
Db	141	TTGGCCATCTGCTAATCGCTGGCTGGCGGATTTCTCTCGCCTCTTCGGCGTCATC	200
QY	41	ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr	60
Db	201	CGTTTCGAGTCGATTATCATGAGTTTGATCCGGTTCAACTACCGGGCCACCGCCTAC	260
QY	61	MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro	80
Db	261	ATGGTGCAGATGGTTGGTACAACTTCTCAACTGGTTCGACGAGCGCATGGTATCCG	320
QY	81	LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle	100
Db	321	CTCGGAGGATGTGGGGGTACCGTCTATCCGGCCTGATGATTACGTCGGCGGAATC	380
QY	101	HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120
Db	381	CATTGGCTCTGCAGTACTCAACATACGGTCCATATTCGTGACATCTGGTGTCTCTG	440
QY	121	AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp	140
Db	441	GGCCGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGACCAAGGAGCTGTGG	500
QY	141	SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer	160
Db	501	TCCGGGGCGCGGCTCTTCGCCGCCAGCTTCATCGGCATCGTGCCTGGCTACATCAGT	560
QY	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180
Db	561	AGTCCGTGGCTGGATCGTACGATAACGAGGGCAATGCCATATTCGCCCTGCAGTTACC	620
QY	181	TyrPheLeuTrpValArgSerValIysThrGly	191
Db	621	TACTTCTGTGGTGGCTCAGTGAAGACTGGA	653

Search completed: December 15, 2004, 06:14:19
 Job time : 6000.52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:51:45 ; Search time 8574.68 Seconds
(without alignments)
4268.640 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 4046

Sequence: 1 MNRTPKMLNSKVAGYSSLLT.....RRKGVIIRNPVWVGKRTLK 774

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131519_8017/app.query.fasta_1.3740
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10028384@cgn_1_1_20362@runat_14122004_131519_8017 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.ste.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4046	100.0	2417	3 AF132552	AF132552 Drosophil
2	4046	100.0	2417	6 AX799088	AX799088 Sequence
3	4046	100.0	2895	6 CQ589353	CQ589353 Sequence
4	3918.5	96.8	4922	6 CQ589352	CQ589352 Sequence

C	5	3918.5	96.8	162921	3	AC007853
C	6	3918.5	96.8	181132	3	AC008206
C	7	3918.5	96.8	227219	3	AE003750
C	8	3864.5	95.5	75650	2	AC018145
	9	2954	73.0	4236	10	BC052433
	10	2949	72.9	4281	6	AX799082
	11	2948	72.9	2481	6	AX074880
	12	2948	72.9	2710	6	AX799084
	13	2585	63.9	54118	3	U13019
	14	2226	55.0	2855	6	CQ573675
	15	2209.5	54.6	2481	5	BC067313
	16	2198.5	54.3	2736	10	BC037612
	17	2198.5	54.3	3094	6	AX799090
	18	2198.5	54.3	3094	10	MUSRPR
	19	2194.5	54.2	3118	9	BT007100
	20	2194.5	54.2	2118	12	BT008132
	21	2194.5	54.2	2450	9	BC048348
	22	2194.5	54.2	2458	6	CQ834714
	23	2194.5	54.2	2472	6	CQ727867
	24	2194.5	54.2	2516	9	BC020965
	25	2193.5	54.2	2293	9	HUMTMC
C	26	2191.5	54.2	110000	8	CR382128_21
	27	2190.5	54.1	3666	10	AK129027
	28	2198	54.1	2608	5	BC048072
	29	2188	54.1	2567	5	BC063234
	30	2187.5	54.1	2898	9	HSN809254
	31	2177.5	53.8	2472	6	AX799092
	32	2177.5	53.8	2472	9	HUMBSA
	33	2176	53.8	2256	6	AX489023
	34	2174	53.7	2435	6	CQ834710
	35	2168.5	53.6	2760	6	CQ492624
	36	2162.5	53.4	110000	8	CR382134_03
	37	2118	52.3	2555	8	BT012881
C	38	2092.5	51.7	110000	8	CR382123_09
C	39	2092.5	51.7	110000	8	CR382123_10
	40	2079.5	51.4	300029	8	AF016903
C	41	2061.5	51.0	40907	8	SPBC1271
	42	2060.5	50.9	2756	8	AX099138
	43	2058	50.9	5048	8	AB015232
	44	2057.5	50.9	2819	8	AB016769
	45	2036.5	50.3	3633	8	SCYGL022W

ALIGNMENTS

RESULT 1	AF132552	AF132552	2417 bp	mRNA	linear	INV 12-FEB-2003
LOCUS	Drosophila melanogaster	Drosophila melanogaster	GM01838	full insert	CDNA.	
DEFINITION	AF132552	AF132552				
ACCESSION	AF132552.1	AF132552.1	GI:4689327			
VERSION	FLI CDNA.					
KEYWORDS	Drosophila melanogaster (fruit fly)					
SOURCE	Drosophila melanogaster					
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
REFERENCE	1 (bases 1 to 2417)					
AUTHORS	Tsang, G., Brokstein, P., Frise, E., Harvey, D., Evans-Holm, M., Lewis, S.E., Suh, C. and Rubin, G.M.					
TITLE	Direct Submission					
JOURNAL	Submitted (02-MAR-1999) Berkeley Drosophila Genome Project, University of California Berkeley, 539 Life Sciences Addition #3200, Berkeley, CA 94720, USA					
REFERENCE	2 (bases 1 to 2417)					
AUTHORS	Stapleton, M., Brokstein, P., Hong, J., Agbayani, A., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Kronmiller, B., Li, P., Liao, G., Miranda, A., Mungall, C.J., Nuno, J., Pacleb, J., Paracas, V., Park, S., Patel, S., Phuanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M. and Ceiniker, S.					
TITLE	Direct Submission					
JOURNAL	Submitted (12-FEB-2003) Berkeley Drosophila Genome Project,					

Lawrence Berkeley National Laboratory, One Cyclotron Road,

Berkeley, CA 94720, USA

COMMENT

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory

Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES

source

1..2417

Location/Qualifiers

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

c1..2417

/gene="OstStt3"

/db_xref="FLYBASE:FBgn0011336"

c1..2395

/gene="OstStt3"

/note="Longest ORF"

/codon_start=2

/product="GM01838p"

/protein_id="BAD27851.2"

/db_xref="GI:28317200"

/db_xref="FLYBASE:FBgn0011336"

/translation="LSECVAFQIGNFRVTVGMNDRITKMLNKKVAGYSSLIITF
 AILLTALAGFSSRLFAVIFESIIIEFDPFNVRATAYMVQNGWYNFLNMFDERAWY
 PLGRIVGTVPLGMITFGGIHLLVNIPIHDIRICVFLAPIFSGLSISTYLLTK
 ELWAGAGLPAASATGIVPGVIRSVAGSDNEGIAIPALQPTFLWVRSVKTVGSFV
 SAAALSVFVMSAWGVFIINLPLRVFLLMGYSRPLLTSYFNFLVGLLAFSM
 QIPVGPQPIRTSEHMAALGVFLLMAVATLHQLSVLSRNEFRKLFIIVGGLLVGV
 EVAVVLLMGVAPWNGRFSYSLWDTGAKIHPIIASVSEHQPTTFWFSFFDLHLV
 CAFPVNYICQINDREVFVLYIASVTFAGVMVRLMLTLTPVVMAGVAFSGLL
 DVFQEDSSKRMGTASAEVEDEEDSEKTLXDKAGLKLKRTKHQAQDQTVSSN
 LKSTIVLAILMLMFAVHCTWTSNAYSSTIVLAFHNSQDSRNILDDFREAYNL
 SQNTADDAVMSMDYGYQIAGMANRTTLVDNNTWNSHIALVGMASSTREKSYEM
 TSLDVIDVLIFFGVIGYGGDINKFLMWVRIAEGBHPKDIKESDYFTDRGFRVDAE
 GAPALLNCLMYKLSYFEGELKLDYSGSPGYDTRNAVIGNKDFDLTVLEAYTTEHM
 LVRIYRVKKPHEFNRLSKERTIIPANFISKNRKRKGYIRNRPVVVKRRLK"

ORIGIN

Alignment Scores:

Pred. No.:	3,14e-314	Length:	2417
Score:	4046.00	Matches:	774
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-10-028-384-8 (1-774) x AF132552 (1-2417)

Db	191	CGTTTCGAGTCGATTATCCATGAGTTTGATCCGTCGTTCACTACCGGGCCACCGCTAC	250
Qy	61	MetValGlnAsnGlyTrpTyrAsnPhelLeuAsnTrpPheAspGluAtcAlaTrpTyrPro	80
Db	251	ATGGTGAGAAATGGTTGGTACAACTCTCTCAACTGGTTCGACGAGCGGCATGGTATCCG	310
Qy	81	LeuGlyVargileValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle	100
Db	311	CTCGGAGGAGTGTGGGGGTACCGTCTATCCGCGCTGATGATTACGTCGCGGGAATC	370
Qy	101	HistTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120
Db	371	CATTCGCTGCTGCACGACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCTG	430
Qy	121	AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp	140
Db	431	GCGCCGATCTTCAGTGGCTACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGTGG	490
Qy	141	SerAlaGlyAlaGlyLeuPheAlaIleAspPheIleAlaIleValProGlyTyrIleSer	160
Db	491	TCCGCGGGCGCGCGCTCTTCGCGCGCAGCTTCATCGCCATCGTCCCTGGGTACATCAGT	550
Qy	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180
Db	551	AGTTCGGTGGCTGATCGTACGATACGATACGAGGSCATTGCCATATTCGCCCTCAGTTCAC	610
Qy	181	TyrPheLeuTrpValArgSerValIlyThrGlySerValPheTrpSerAlaAlaAla	200
Db	611	TACTTCTGTGGTGGCTCGCTCAGTGAAGACTGATCCGCTGCTGGTGGCGCGACCGCT	670
Qy	201	LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle	220
Db	671	TTGTCTCTCTTCTACATGGTGTCCGCTGGGTGGCTACGTTGTTTCATCATCAACCTGATA	730
Qy	221	ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer	240
Db	731	CCCTCGACGTCTCTGATCTGCTCATTAATGGACAGTACTCGCGCGTCTGCTCACCAGC	790
Qy	241	TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe	260
Db	791	TACAGCACCTTCTACATCTCGGACTGCTGTTTCTCCATGCAGATCCCTCTCGTGGGATTC	850
Qy	261	GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla	280
Db	851	CAACCATACGACACAGTGAACACATGCTGCGCTGGCGTGGGAGTGTGTGCTCTTATGCC	910
Qy	281	ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe	300
Db	911	GTGGCCACCTTGGCCCATTTTCAGTCCGCTGCTGTCGCGCAACGAGTTCGCGAAGCTGTT	970
Qy	301	IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuLeuThrMet	320
Db	971	ATCGTCGGCGGAGTGTGCTGGTGGCGCTGGCGCTTGTGGCGCGTGGTGGTCTCACCATG	1030
Qy	321	LeuGlyValValAlaAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla	340
Db	1031	CTGGCGTGTGTGGCGCGGTGGAGTGGAGCTTCTACTCGTGTGGGATCTGGGTACGCC	1090
Qy	341	LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer	360
Db	1091	AAGATCCACATTCCTCATTCATCGCTGGTGGAGCATCAGCCACCATCTGTTCTTCG	1150
Qy	361	PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle	380
Db	1151	TTCTTCTTTGATCTGCACATCTGCTGTGGCGCTTCCCGAGTGGGAGTGTGGTACTGCATC	1210
Qy	381	LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe	400
Db	1211	AAGCATCAACGACGAGCGCTTTTCGTGGTGGTGTGTACGCCCATCAGTGGCGTTTACTTC	1270
Qy	401	AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly	420
Db	1271	GCTGGTGTGATGGTGGTGTGATGTGACCTCACCGCTGGTGGTGTGTCATCTGGCGGA	1330

QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 DB 1331 GTGGCTTTTCGGGACTGTGGATGTTCTCTGCAAGAGGATTCTGTCTAAGCGAATGGC 1390
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLeuLysThr 460
 DB 1391 ACAGCCATTAAGCGCCACCCAGAGTGGATGAGCTGAGGATTCATITGAGAGAGACG 1450
 QY 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThr 480
 DB 1451 CTGTACGCAAGCGCTGGCAAGCTGAAGCATGCTACTAAGCATGATGCCAGCAGATACT 1510
 QY 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet 500
 DB 1511 GGCGTCAGCTCCAACTGAGAGATATGTTATTTTGGCCGTTCTTAATGCTTGTATGATG 1570
 QY 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
 DB 1571 TTGCTGTCTCACTGCACGTGGGTGACCAAGCATGCTACTCCAGTCCCTCCATGTCTTG 1630
 QY 521 AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyr 540
 DB 1631 GCTTTCCAAACAGTCAGATGATCCCGCAACATTTTAGACGATTTTCAGAGAGCTTAC 1690
 QY 541 TyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGly 560
 DB 1691 TACTGCTTTTCGACAGAACACTCCGATGATGCTCGCTTATGTTCTTGTGGGATTACGGA 1750
 QY 561 TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 580
 DB 1751 TACCAGATAGCGGGAATGCAACAGACAGACGCTAGTGGATATATATACGTGGACAT 1810
 QY 581 SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle 600
 DB 1811 AGTCACATAGCGCTGTTGGCAAGCAATGTCTTCAACCGAGGAGAGTCTTACGAAAT 1870
 QY 601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620
 DB 1871 ATGACATCTCTTGACGTGACATGCTTTTGGTGATCTTTGGCGGTGTGATCGGCTATCT 1930
 QY 621 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 640
 DB 1931 GSCGATGATCAACAAAGTCTCTGTGGATGTGTCCGATTTGCTAGGAGAGACATCCCAAG 1990
 QY 641 AsnIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly 660
 DB 1991 GACATTAAGGAAGCGATTTACTTTTACCAGCCGCGTGAAATTCAGGCTAGATGCCAAGGT 2050
 QY 661 AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680
 DB 2051 GCTCCGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGGAATTG 2110
 QY 681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
 DB 2111 AGTTGGACTACAGAGTCCATCTGGATGATGATGCACACGTAAACCGCTCATTTGGGAAT 2170
 QY 701 LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArg 720
 DB 2171 AAGGACTTCGATCTGACCTACCTGAGGAGGCGCTACACACAGAACACTGCTGCTTGTTCG 2230
 QY 721 IleTyrArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
 DB 2231 ATCTATAGGTGAAGAGCGGATGATGATTTCAATAGACCATCACTGAAGACCAAGAGAGA 2290
 QY 741 ThrIleProAlaAsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIle 760
 DB 2291 ACGATTCCTCCAGCAAACTTCAATTCGAGAAAGAACTCTAAGCGTCCGAGGGCTACATA 2350
 QY 761 ArgAsnArgProValValLysGlyLysArgThrLeuLys 774
 DB 2351 CGAAACCGACCGCTGCTTTTGAAGGGAAGAAACCACTTGAAA 2392

RESULT 2
 AX799088
 LOCUS 2417 bp DNA linear PAT 08-OCT-2003
 DEFINITION Sequence 7 from Patent WO03054008.
 AX799088
 ACCESSION AX799088
 VERSION AX799088.1 GI:37605061
 KEYWORDS Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1
 AUTHORS Perreault, C. and McBride, K.
 TITLE Mammalian SIMP protein, gene sequence and uses thereof in cancer
 therapy
 JOURNAL Patent: WO 03054008-A 7 03-JUL-2003;
 COMPATIGENE inc. (CA)
 FEATURES
 LOCATION/Qualifiers
 1..2417
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /note="SIMP, Acc# AF132552"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,148-314 Length: 2417
 Score: 4046.00 Matches: 774
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-8 (1-774) x AX799088 (1-2417)
 QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 DB 71 ATGATTCGAGCGCGAAGATGCTGAACAGCAGAGTGGTGGTACACAGCCTAATCACC 130
 QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 DB 131 TTCCGCATCTCTGTAATCGCTGGTGGCGGATTTTCCTCTCGCTCTCTCGCGTCTATC 190
 QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
 DB 191 CGTTTCGAGTCGATATCCATGAGTTGATCCGCTGGTTCAACTACCGGCGCACCGCTAC 250
 QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
 DB 251 ATGGTGCAGAAATGTTGTGTACAACTTCTCAACTGGTTTCGACGAGCGCATGGTATCCG 310
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 DB 311 CTCGCGAGATTTGGCGGTACCGTCTATCCCGCTGATGATTAGTCCGCGGATC 370
 QY 101 HistTrpLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 DB 371 CATTCGCTGCTGACGTACTCAACATACCGGTCCATATTCGTGACATCTGCGTGTCTG 430
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 DB 431 GGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGTGG 490
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 DB 491 TCCCGGCGCGCGGCTCTTCGCGCCAGCTTCATCGCCATCGTGTGCTGCTATCAGT 550
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 DB 551 AGTTCGTTGGCTGGATCTGTACAGAGAGGCAATTCGCATATTCGCGCTGCGGTTCCAC 610
 QY 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheThrSerAlaAlaAla 200

611 TACTTCCTGGTGGCTCAGTGAAGTGGATCGTGTCTGTGGTGGCGGAGCGCT 670
 201 LeuSerThrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleLeuIle 220
 671 TTGTCCTACTTCTCATGTGTGTCGGCTGGGGTGGCTACGTGTGTCTCATCAACCTGATA 730
 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
 731 CCCCTGCAAGCTTCTGACTGCTCATATGGCAGAGTACTCGCGGGTGTGTGACCCAGC 790
 241 TyrSerThrPheTyrIleLeuGlyLeuPheSerMetGlnIleProPheValGlyPhe 260
 791 TACAGACCTTCTCATCTCGGACTGCTGTCTCCATGAGATCCCTCTCGTGGGATTC 850
 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 851 CAACCGATACGACACGACGATGAACATGCTGGCTGGAGTGTGTGTCTCTATATGGCC 910
 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGlnPheArgLysLeuPhe 300
 911 GTGGCCACCTTGGCCCATTTGCACTCGTGTCTCGCGCAACGAGTTCGCGAGCTGTTC 970
 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 971 ATCGTGGCGGATTTGCTGTGGTGGGGTGGGCTCTTTGTGGCGCTGTGTGTCTACCATG 1030
 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 1031 CTGGCGCTGTGGCCCGCTGGAGTGGAGCTTCTACTCGTGTGGGATATCGCTACGCC 1090
 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrTrpPheSer 360
 1091 AAGATCCACATTCCTCATCTGATCGTGTGGAGCATCAGCCACCTTGGTCTCG 1150
 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
 1151 TTCCTCTTTGTATCTGCATCTCGTGTGGCTTCCAGTGGAGTGTGGTACTGCATC 1210
 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 1211 AAGCAGATCAACGACGAGCGCGTTCGTGTGTCTGTACGCCATCAGTGGGTACTTC 1270
 401 AlaGlyValMetValArgLeuMetLeuThrIleuThrProValValCysMetIleAlaGly 420
 1271 GCTGTGTGTGTGGTGGCTGTGTGATGTTGACCTCACCGCGTGTGTGTGTGTGGCGGA 1330
 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 1331 GTGGCTTTTTCGGGACTGTTGGATGTGTTCTCGAAGAGATTCGTCCTAAGCGAATGGGC 1390
 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThr 460
 1391 ACAGCCATAAGCGAGCGACCGACGAGTGGATGCAAGCTGAGGATTCATTCAGAGAGAGC 1450
 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThr 480
 1451 CTGTACGACAAGCTGGCAAGCTGAACATCGTACTAAGCATGATCCCGCAGGATACT 1510
 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuMetMet 500
 1511 GCGGTACGCTCAACCTCAAGAGATTTGTTATTTTGGCGCTTCTAATGCTGTGTGATG 1570
 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
 1571 TTCGTGTCTCCACTGACGTGGTGGTGAACGACCAATGCTCTACTCCAGTCCCTCCATGTCITG 1630
 521 AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyr 540
 1631 GCTTTCACCAACAGTCAAGATGGATCCGCAACATTTAGACGATTCAGAGAGGCTTAC 1690
 541 TyrTrpLeuSerGlnAsnThrAlaAspAlaArgValMetSerTrpTrpAspTyrGly 560

Db 1691 TACTGGCTTTCCGAGAACACTGCCGATGATGCTCGCTTATGTCTTGTGGGATTACGGA 1750
 Qy 561 TyGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 580
 Db 1751 TACCAGATAGCGGGAATGGCAACAGACGAGCTAGTGGGATAATAATACGTGGAAACA 1810
 Qy 581 SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle 600
 Db 1811 AGTCACATAGCGCTGTGTGGCAAGCAATGTCTTCAACCGAGGAGAGTCTCTACGAAAT 1870
 Qy 601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620
 Db 1871 ATGACATCTCTGACGTGACTACCTTTTGGTGATCTTTGGCGGTGATCGGTATCTCT 1930
 Qy 621 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 640
 Db 1931 GGCAGATATCAACAAGTTCCTGTGGATGGTCCGAAATTCGTGAGGAGAGACATCCCAAG 1990
 Qy 641 AspIleIleGlySerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGlyGly 660
 Db 1991 GACATTAAAGGAAGCGATTAATTTACCGACCGCGGTGAATTCAGGGTAGATGCCGAAGT 2050
 Qy 661 AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680
 Db 2051 GCTCCGCGCTCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGAATTG 2110
 Qy 681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
 Db 2111 AAGTTGGATCAGAGTCCATCTGGATATGATGCACACGACGCGCTCATTTGGGAAT 2170
 Qy 701 LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArg 720
 Db 2171 AAGGACTTCGATCTGACCTTACCTGGAGGAGGCTTACACACAGAACACTGCTGTTGTCG 2230
 Qy 721 IleTyrArgValLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
 Db 2231 ATCTATAGGTGAAGAGCGCATGAGTTCATATAGACCATCCTCAAGAGACCAAGAGAGA 2290
 Qy 741 ThrIleProAlaAsnPheIleSerArgLysAsnSerLysArgLysGlyTyrIle 760
 Db 2291 ACCATTCTCCAGAACTTCATTTGAGAAAGAACTCTAAGCGTCCGCAAGGCTACATA 2350
 Qy 761 ArgAsnArgProValValLysGlyLysArgThrLeuLys 774
 Db 2351 CGAACCAGCCCGTGTGTGTTAAGGAAACGAACTTGAAT 2392

RESULT 3
 LOCUS CQ589353 2699 bp DNA linear PAT 02-FEB-2004
 DEFINITION Sequence 17111 from Patent WO0171042.
 ACCESSION CQ589353
 VERSION CQ589353.1 GI:41648215
 KEYWORDS
 SOURCE Drosophila sp.
 ORGANISM Drosophila sp.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
 AUTHORS Detection kits, such as nucleic acid arrays, for detecting the
 TITLE expression of 10,000 or more Drosophila genes and uses thereof
 JOURNAL Patent: WO 0171042-A 17111 27-SEP-2001;
 PE Corporation (NY) (US)
 FEATURES Location/Qualifiers
 source 1..2699
 /organism="Drosophila sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:7242"

ORIGIN
 Alignment Scores: 3.62e-314 Length: 2699
 Pred. No.:

Score: 4046.00 Matches: 774
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-028-384-8 (1-774) x CQ589353 (1-2699)

QY	1	MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr	20
DB	80	ATGAATCGGACCCGCAAGATGCTGAACAGCAAGGTGGCTGCTACAGCAGCCTAATCACC	139
QY	21	PheAlaIleLeuLeuIleAlaThrLeuAlaGlyPheSerArgLeuPheAlaValIle	40
DB	140	TTGCGCATCTCTGCTAATCGCTGGCTGGCCGGATTTCTCTCGCCCTCTGCGCGTATC	199
QY	41	ArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrA:GAlaThrAlaTyr	60
DB	200	CGTTTCGAGTCGATATATCATGAGTTTGATCCGTGGTTCAACTACCGGGCCACCGCCTAC	259
QY	61	MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluAlaTyrPro	80
DB	260	ATGGTGCAAGATGGTGTGTAACAATCTCTCAACTGGTTCGACGAGCGCATGTATCCG	319
QY	81	LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle	100
DB	320	CTCGGAGATTTGGGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGGCGGAATC	379
QY	101	HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120
DB	380	CATTGGGTGTCGACGTACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCTG	439
QY	121	AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp	140
DB	440	GGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGTCGCAAGGAGCTGTG	499
QY	141	SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer	160
DB	500	TCCCGGGCGCGGCTCTTCGCGCGCAGCTTCATCGCCATCTGCTGCTGCTGCTACATCAGT	559
QY	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180
DB	560	AGTTCGGTGGTGGATCGTACGATACAGGGGCAATTCGCCATATTCGCCCTGCAAGTTCACC	619
QY	181	TyrPheLeuTrpValArgSerValLysThrGlySerValPheThrSerAlaAlaAlaAla	200
DB	620	TACTTCCTGCTGGGTGGCTCAGTGAAGACTGGATCCGTGTTCTGTGTCGGCGGAGCGGT	679
QY	201	LeuSerTyrPheThrMetValSerAlaTrpGlyGlyTyrValPheIleLeuLeuIle	220
DB	680	TTGTCCTACTTCTACATGGTGTCCGCTGGGGTGGCTACGTGTTTCATCATCAACCTGATA	739
QY	221	ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer	240
DB	740	CCCTCGCACCTCTTCGTACTGCTCATATGGGAGAGTATCTCGCGGCTGTGCTGACCAAG	799
QY	241	TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe	260
DB	800	TACAGCACCTTACATCTCTGGACTGCTGTTCTCCATGCAAGTCCCTTCGTGGGATTC	859
QY	261	GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla	280
DB	860	CAACCATACGACCAAGTGAACATGGCTGGCTGGGAGTGTGTGTCTCTTATGGCC	919
QY	281	ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe	300
DB	920	GTGGCCACCTTGGCCATTTGCACTCCGTGCTGTCGCAACAGAGTTCGCGAGCTGTTC	979
QY	301	IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet	320
DB	980	ATCGTCGGCGGATGCTGGTGGCGGTGGCGTCTTTGTGGCGCTGCTGCTGCCATG	1039
QY	321	LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla	340

DB	1040	CTGGGGGTGTGGCCCGCTGGAGTGAGCGCTTCTACTCGTGGGATATGCTGCTACGGC	1099
QY	341	LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer	360
DB	1100	AAGATCCATATCCCATCATTCATCCGTGTCGAGCATCAGCCACCATCTGGTTCTCG	1159
QY	361	PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle	380
DB	1160	TTCTTTCTTGTATCTGCACATCTCTGGTGGCTTCCAGTGGAGTGTGGTACTGCATC	1219
QY	381	LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe	400
DB	1220	AAGCAGATCAACAGCAGCGCGTTTTCGTGGTGTGTACGCCATCAGTGGCGGTTCATC	1279
QY	401	AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly	420
DB	1280	GCTGGTGTGATGGTGGTTGATGTGACCTCAGCCGGTGTGTGATGTGTCGCCGGA	1339
QY	421	ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly	440
DB	1340	GTGGCCTTTTCGGGACTGTGGATGTGTTCTCTGCAAGAGGATTCCTTAAGCGAATGGC	1399
QY	441	ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr	460
DB	1400	ACAGCCATTAAGCGCAGCCACCGAAGTGAAGCTGAGGATTCATTCAGAAAGACG	1459
QY	461	LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThr	480
DB	1460	CTGTACGACAAAGCTGGCAGCTGAAGCATCGTACTAAGCATGATGCCAGCAGGATACT	1519
QY	481	GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuMetMet	500
DB	1520	GGCGTCAGCTCCAACTGAAGAGTATTGTTATTTGGCCGTTCTAATGCTGTGTGATG	1579
QY	501	PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu	520
DB	1580	TTGCGTGTCCACTGCGCTGGTGACCAAGTGCCTACTCCAGTCCCTCCATGTCGTG	1639
QY	521	AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyr	540
DB	1640	GCTTTCCCAACAGTCAAGATGGATCCCGCAACATTTTAGACGATTTTCAGAGAGCTTAC	1699
QY	541	TyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGly	560
DB	1700	TACTGGCTTTCGCAACACTGCGCATGATGCTCGCGTATGCTGTGGTGGATTCGGA	1759
QY	561	TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn	580
DB	1760	TACCAGATAGCGGGAATGGCAACAGACGCTAGTGGATATATATACGTGGAACAT	1819
QY	581	SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle	600
DB	1820	AGTCACATAGCGCTGGTTGGCAAGCAATGTTCAACCGGAGAGAGTCTACGAAAT	1879
QY	601	MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer	620
DB	1880	ATGACATCTCTGACGTGGACTACGTTTGGTGAATCTTGGCGGTGATCGGCTATCT	1939
QY	621	GlyAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys	640
DB	1940	GGCGATGATATCAACAAGTTCCTGTGGATGTCGGAATTCGTAGGGAGAGATCCCAAG	1999
QY	641	AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly	660
DB	2000	GACATTAAGNAACGATTTACTTTACCGACCGCGTGAATTCAGGGTAGATGCCAAGGT	2059
QY	661	AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu	680
DB	2060	GCTCCGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGGAATG	2119
QY	681	LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn	700

```

Db      2120  AAGTTGGACTACAGAGGTCCTCGATATGATCGCACACGTAACGCGCTCATTTGGGAAT 2179
QY      701  LysAspPheAspLeuThrTyrluGluGluAlaTyrluThrGluHisTrpLeuValArg 720
Db      2180  AAGGACTTCGATCTCACTTACCTGAGGAGGCTTACACACAGAACACTGGCTTGTTCGC 2239
QY      721  IleTyArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
Db      2240  ATCTATAGGTGAGAGCGCATGAGTTCAATAGACCATCACTGAAGACCAAGGAGAGA 2299
QY      741  ThrIleProAlaAsnPhelIleSerArgLysAsnSerLysArgArgLysGlyTyrlle 760
Db      2300  ACGATTCTCCAGCAAACTTCTTCGAGAAAGAACTCTAAGCGTCGCAAGGGCTACATA 2359
QY      761  ArgAsnArgProValValLysGlyLysArgThrLeuLys 774
Db      2360  CGAAACCGACCGGTGTGTTTAAAGGAAACGAAACCTTGAAA 2401

RESULT 4
CQ589352/c      4922 bp      DNA      linear      PAT 02-FEB-2004
LOCUS      Sequence 17110 from Patent WO0171042.
DEFINITION      CQ589352
ACCESSION      CQ589352
VERSION      CQ589352.1 GI:41648214
KEYWORDS
SOURCE
ORGANISM      Drosophila sp.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
Detection kits, such as nucleic acid arrays, for detecting the
expression of 10,000 or more Drosophila genes and uses thereof
Patent: WO 0171042-A 17110 27-SEP-2001;
PE Corporation (NY) (US)
FEATURES
source      1..4922
            Location/Qualifiers
            /organism="Drosophila sp."
            /mol_type="unassigned DNA"
            /db_xref="taxon:7242"

ORIGIN
Alignment Scores:      1.28e-303      Length:      4922
Pred. No.:      3948.50      Matches:      772
Score:      90.93%      Conservative:      0
Percent Similarity:      90.93%      Mismatches:      2
Best Local Similarity:      96.85%      Indels:      77
Query Match:      6      Gaps:      4
DB:

US-10-028-384-8 (1-774) x CQ589352 (1-4922)
QY      1  MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrlSerSerLeulleThr 20
Db      3843  ATGAATCGGACGCGCAAGATGCTGAACAGCAAGGTGGCTGGCTACAGCAAGCTTAATCAC 3784
QY      21  PheAlaIleLeulleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db      3783  TTCGCCATCTGCTAATCGCTGGCTGGCGGATTTCTCTCGCTCTTCGCGCTATC 3724
QY      41  ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrAlaTyrl 60
Db      3723  CGTTTCGAGTCGATTAATCCATGAGTTTGATCCGCTGGTTCAACTACCGGGCCACCGCTAC 3664
QY      61  MetValGlnAnGlyTyrlTrpAsnPhelIleAsnTrpPheAspGluArgAlaTrpTyrlPro 80
Db      3663  ATGGTGCAAGATGTTGTACAACTTCTCACTGGTTCCAGCGCGCGCATGGTATCCG 3604
QY      81  LeuGlyArgIleValGlyGlyThrValTyrlProGlyLeuMetIleThrSerGlyGlyIle 100
Db      3603  CTCGGCAGGATTTGGGGCGGTACCGCTATATCCCGCGCTGATGATTACGTCGCGGATC 3544

```

```

QY      101  HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db      3543  CATTGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCTG 3484
QY      121  AlaProIlePheSerGlyLeuThrSerIleSerThrTyrlLeuLeuThrLysGluLeuTrp 140
Db      3483  GGGCGGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGACCAAGAGCTGTGG 3424
QY      141  SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrlIleSer 160
Db      3423  TCCGCGGGCGCGGCTCTTTCGCCCGCAGCTTCATCGCCATCGTGCCTGGCTACATCAGT 3364
QY      161  ArgSerValAlaGlySerTyrlAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db      3363  AGTCCGCTGGCTGGATCGTACGATTAACGAGGCGATTCGATATTCGCCCTCGAGTTCAC 3304
QY      181  TyrlPheLeuTrpValArgSerValLysThrGlySerValPheThrPheTrpSerAlaAla 200
Db      3303  TACTTCTGTGGTGGCTCAGTGAAGACTGATCCGTGTTCTGTGCGCGCGCAGCCGCT 3244
QY      201  LeuSerTyrlPheTyrlMetValSerAlaTrpGlyTyrlValPheIleIleAsnLeulle 220
Db      3243  TTGTCCTACTTCTACATGTTCCGCTGGGTGGCTAGCTGTTTATCATCACTCACTGATA 3184
QY      221  ProLeuHisValPheValLeuLeulleMetGlyArgTyrlSerProArgLeuLeuThrSer 240
Db      3183  CCCTCGACGCTCTTCTGACTGCTCATTAATGGCGAGGTACTCGCCGCTGCTGACCCAGC 3124
QY      241  TyrlSerThrPheTyrlIleLeuGlyLeuPheSerMetGlnIleProPheValGlyPhe 260
Db      3123  TACAGCACCTTCTACATCTCGGACCTGCTTCTCCATGCATCCCTTCTGCGGATTC 3064
QY      261  GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
Db      3063  CAACCGATAGCACCACTGACACATGGCTGGCTGGAGTGTGTTGCTCCTTATGGCC 3004
QY      281  ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
Db      3003  GTGGCCACCTTGGCGCAATTTGCAGTCCGCTGTGTCGCCAAGAGTTCGGAAGCTGTTC 2944
QY      301  IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
Db      2943  ATGTCGCGCGGATTTGCTGGTGGCGCTTGGCTGCTTGTGGCGCTGCTGCTCACCAGT 2884
QY      321  LeuGlyValValAlaProTrpSerGlyArgPheTyrlSerLeuTrpAspThrGlyTyrlAla 340
Db      2883  CTGGCGCTTGTGGCGCGCTGGAGTGGACGCTTCTACTCGCTGGGATACTGGCTACGCG 2824
QY      341  LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
Db      2823  AAGATCCACATTCCTCATTCGCTGTCGAGCATCAGCCACCCACCTTGGTTCGTG 2764
QY      361  PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrlCysIle 380
Db      2763  TTCTTCTTGTATCTGCACATCTCTGGTGGCGCTTCCAGTGGGAGTGTGTTATCGATC 2704
QY      381  LysGlnIleAsnAspGluArgValPheValValLeuTyrlAlaIleSerAlaValTyrlPhe 400
Db      2703  AAGCAGATCAACGACGAGCGCTTTCGTGGTGTGTGACGCCATCAGTGGCGTTTACTTC 2644
QY      401  AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
Db      2643  CTTGGTGTGATGGTGGCTTGTATGTTGATCTCACCTCACGCGCGTGGTGTGATGCTGG 2584
QY      421  ValAlaPheSerGlyLeuLeuAspValPheLeuGlnLysAspSerSerLysArgMetGly 440
Db      2583  GTGGCGCTTTCGGGACCTTGTGGATGTTCTCTCGAAGAGATTCGTTAAGCCGATGGCG 2524
QY      441  ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
Db      2523  ACAGCCATACGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTCATTTGAGAAGACG 2464
QY      461  LeuTyrlAspLysAla----- 465

```

```

Db      2463 CTGTACGACAGGT-GAGTTCTTACTAACACACATCCATCGGATTTGATTTTAAATAACA 2405
Qy      466 -----GlyValLeuLeuHisArgThrLysHisAspAlaGlnGlnAspThr 480
Db      2404 TCGCATTTTGCAGGCTGGCAAGCTGAAGCATGTAAGCATGATGATCCCGACGAGGATCT 2345
Qy      481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet 500
Db      2344 GCGGFCAGCTCAACCTGAAGAGATTGTTATTTTGGCCGCTTCTAATGCTGTGTGATGATG 2285
Qy      501 PheAlaValHisCysThrTyrValThrSerAsnAlaTyrSerProSerIleValLeu 520
Db      2284 TTTCGTGTCCACTGCACGTGGTGCACGCAATGCTTACTCCAGTCCCTCCATTGCTTG 2225
Qy      521 AlaPheHisAsnSerGlnAspGly----- 528
Db      2224 GCTTTTCCACACAGTCAAGATGGGTAAAGCAAGTAGCATCATGCTTCTTTAAACCTTA 2165
Qy      529 -----SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyr 541
Db      2164 TCTAACCTTTTCTCCACTAGATCCGCAACATTTTAGCAGATTTCAGAGAGGCTTACTAC 2105
Qy      542 TrpLeuSerGlnAsnThrAlaAspAlaArgValMetSerTyrTyrAspTyrGlyTyr 561
Db      2104 TGGCTTTTCCGAGAACACTGCGGATGATGCTCGCTTATGCTTGGTGGGATTACGGATAC 2045
Qy      562 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTyrAsnAsnSer 581
Db      2044 CAGATAGCGGAATGGCAACAGAACGACGCTAGTGGTAATATACGTCGGAACATAGT 1985
Qy      582 HisIleAlaLeuValGlyValAlaMetSerSerThrGluGluLysSerTyrGluIleMet 601
Db      1984 CACATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTACGAAATATG 1925
Qy      602 ThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 621
Db      1924 ACATCTCTTGCAGTGCATACGTTTGTGTGATCTTTGCGGTGTGATCGCTATTCTGGC 1865
Qy      622 AspAspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAsp 641
Db      1864 GATGATATCAACAGTTCTCTGTGATGCTGCGAATTCGTGAGGAGAGCATCCCAAGGAC 1805
Qy      642 IleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAla 661
Db      1804 ATTAAGGAAGCGATTACTTTACCGACCGCGGTGATTCAGGCTAGATCGGAAGTGCT 1745
Qy      662 ProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLys 681
Db      1744 CCGGCCCTGCTCACTGCTTATGTACAAATTAACTACTACAGATTTCGGGGAAATGAAG 1685
Qy      682 LeuAspTyrArg----- 685
Db      1684 TTGGACTACAG-GTAAGCGGAAACATTTCTCAGTAGGATGAGGATGACACACTAAGTGGCTGT 1626
Qy      686 -----GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
Db      1625 TTTCAGAGTCCATCTCGATATGATGCGCACGTAAACCGCTCATTTGGGAATTAAGGACTTC 1566
Qy      704 AspLeuThrTyrLeuGluGluAlaTyrThrGluHisThrLeuValArgIleTyrArg 723
Db      1565 GATCTGACCTACTCGAGGAGGCGCTACACACAGACACTGGCTTGTTCGATCATATAGG 1506
Qy      724 ValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIlePro 743
Db      1505 GTGAAGAGCCGCGATGAGTTCAATAGACCATCTAGAGCAACCAAGGAGAGAGCATTCCT 1446
Qy      744 ProAlaAsnPheIleSerArgLys----- 751
Db      1445 CCAGCAAACTTCATTTCCGAAAGAGTAGTGTAACCTGCAGCTCCCAATGAATTCATTAA 1386
Qy      752 -----AsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765

```

1385 CAATCTGATTTCTTGACAACTCTTAAGCGTCCGAGGGCTACATACGAAACCGACGGTT 1326
 766 ValVallysglyLysArgThrLeuLys 774
 1325 GTTGTAAAGGAAAACGACCTTGAAA 1299

RESULT 5
 AC007853/c 162921 bp DNA linear INV 27-FEB-2001
 LOCUS Drosophila melanogaster, chromosome 3R, region 96B-96C, BAC clone
 DEFINITION BACR03L02, complete sequence.
 AC007853
 VERSION AC007853.4 GI:13129410
 KEYWORDS HTG.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 162921)
 REFERENCE
 AUTHORS
 Holt, R.A., Evans, C.A., Gocayne, J.D., Adams, M.D., Krontmiller, B., Tyler, D., Wan, K.H.,
 Rogers, Y., An, H., Baldwin, D., Banzone, J., Beeson, K.V., Busam, D.A.,
 Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,
 Dodson, K., Dorsett, V., Doupl, E., Doyle, C., Dresnek, D., Farfan, D.,
 Ferris, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
 Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
 Ibegwam, C., Jalali, M., Kruse, D., Li, P., Matti, B., Moshrefi, A.,
 McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nuncio, J.,
 Pacle, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
 Phoenavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
 Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
 Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
 Sequencing of Drosophila chromosome 3R, region 96B-96C
 Unpublished
 2 (bases 1 to 162921)
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Butenik, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Celenik, S.E., Chavez, C., Chew, M., Ciesiolka, L.,
 Butenik, S.E., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Humnasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.
 Direct Submission
 Submitted (17-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Feb 27, 2001 this sequence version replaced gi:5670481.
 COMMENT
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdgp@fruitfly.berkeley.edu.
 FEATURES
 1..162921
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /map="96B-96C"
 /clone="BACR03L02 (D766)"
 /clone_lib="RPCI-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial EcoRI in
 pBACE3.6)"

ORIGIN

Alignment Scores:

Pred. No.: 1,16e-301 Length: 162921
 Scores: 3918.50 Matches: 772
 Percent Similarity: 90.93% Conservative: 0
 Best Local Similarity: 90.93% Mismatches: 2
 Query Match: 96.85% Indels: 77
 DB: 3 Gaps: 4

US-10-028-384-8 (1-774) x AC007853 (1-162921)

QY	1	MetAenArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr	20
DB	108481	ATGAATCGACCGCCAGAGTGTGAACAGACAGAGTGGCTACAGACGCTAATACCC	108422
QY	21	PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValile	40
DB	108421	ITCGCCATCCGCTAATCGCGCTGGCTGGCCGGATTTCCTCTCGCCCTCTCGCCGCTATC	108362
QY	41	ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr	60
DB	108361	CGTTTCAGTGCATTATCCATAGATTGTATCCGTGGTTCACTACCGGCCACCCGCTAC	108302
QY	61	MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro	80
DB	108301	ATGGTGAGAATGGTTGGTCAACTTCCTCAACTGGTTCGACGAGCGGCGATGGTATCG	108242
QY	81	LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyIle	100
DB	108241	CTCGCAGGATTGTGGCGGTACCGTCTATCCCGGCTGATATTACGTCCGCGGGAATC	108182
QY	101	HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu	120
DB	108181	CATTGGGTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATCTGCTGTCTCG	108122
QY	121	AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp	140
DB	108121	GGCGGATCTTCAGTGGCCCTGACCTCCATCTCCACCTACTCTGACCAAGAGGTGG	108062
QY	141	SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer	160
DB	108061	TCCGCGCGCGCGGCTCTTCGCGCGCGAGCTTCATCGCCATCGTGGCTGGCTACATCAGT	108002
QY	161	ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr	180
DB	108001	AGGTGGGTGGTGGATCGTACGATAACGAGGCGCATTCGCCATATTCGCCCTGCGGTACCC	107942
QY	181	TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAlaAla	200
DB	107941	TACTTCCTGTGGGTGGCTCAGTGAAGACTGGATCGTGTCTGTGGCGCGCGAGCGCT	107882
QY	201	LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle	220
DB	107881	TTGTCTTCTTACATGCTGCGCTGGCTGGGTGGCTAGCTGTTTCATCATCACTGATA	107822
QY	221	ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer	240
DB	107821	CCCCGTGACGCTCTTCGTACTGCTCATTATGGGCGAGTACTCGCCGCTGCTGACACAGC	107762
QY	241	TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe	260
DB	107761	TACAGCACCTTCTACATCTGGAGCTGCTGTTCTCCATCGATCCCTTGTGGGATTC	107702
QY	261	GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla	280
DB	107701	CAACCGATACGCACAGTGAACATCGCTGGCTGGCTGGGAGTGTGTGTCTCTATGGCC	107642
QY	281	ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe	300
DB	107641	GTGGCCACCTTGGGCCATTTCGAGTCCGTGCTGTCCGCAACAGATTTCGGAGCTGTTC	107582
QY	301	IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet	320
DB	107581	ATCGTCGGCGGATTGCTGGTGGCGGTGGCGTCTTTGTGGCGCTGCTGTCTCACCATG	107522
QY	321	LeuGlyValValAlaProTrpSerClyArgPheTyrSerLeuTrpAspThrGlyTyrAla	340
DB	107521	CTGGCGCTGTGGCCCGCGGAGTGAGCTTCTACTCGCTGTGGGATCTGGCTACGCC	107462
QY	341	LysIleHisIleProIleAlaSerValSerGluHisGlnProThrTrpPheSer	360
DB	107461	AGATCCACATTCCTCATTCATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG	107402
QY	361	PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle	380
DB	107401	TTCTTCTTTGATCTGCACATCTCTGTGTGGCTTCCAGTGGAGTGTGGTACTGCATC	107342
QY	381	LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe	400
DB	107341	AAGCAGATCAACGACGAGCGGCTTTTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	107282
QY	401	AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly	420
DB	107281	GCTGGGT	107222
QY	421	ValAlaPheSerGlyLeuLeuAspValPheLeuGlnAspSerSerLysArgMetGly	440
DB	107221	GTGGCTTTTCGGACTGT	107162
QY	441	ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr	460
DB	107161	ACAGCCATRAGCGACGACCCGAGTGGATGAAGTGGATTCATTCATTCGAGAGAGACG	107102
QY	461	LeuTyrAspLysAla	480
DB	107101	CTGTACGACAGGT-GAGTTCTTACTACACACATCCATCCGATTTGATTTTAAATACA	107043
QY	466	-----GlyValLeuLysHisArgThrLysHisAspAlaGlnAspThr	480
DB	107042	TGCAATTCGAGCTGGGAGCTGAGCATCTGATGATGATGATGATGATGATGATGATGATG	106983
QY	481	GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet	500
DB	106982	GGCGTCAGCTCCAACTGAAGATATTGTTATTTTGGCCGTTCTTAATGCTGTGTGATG	106923
QY	501	PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu	520
DB	106922	TTGCTGTCTCATCTGACCTGGGTGACCAAGCAATCCCTACTCCAGTCCCTCCATGCTTG	106863
QY	521	AlaPheHisAsnSerGlnAspGly	540
DB	106862	GCTTTCACACACAGTCAAGATGGGTGAAGCAAGTAGCATCATGCTCTCTTTAAACCTA	106803
QY	529	-----SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyr	541
DB	106802	TCTAACCTTTTCTCCACTAGATCCCGCAACATTTTAGACATTTACAGAGGCTTACTAC	106743
QY	542	TrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGlyTyr	561
DB	106742	TGGCTTTCCGACAACTGCGGATGATGCTCGCTTATGTCTTGTGGGATTTACGGATAC	106683
QY	562	GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer	581
DB	106682	CAGATAGCGGATGCGCAACAGACGCGTAGTGGATTAATATACCTGGACATATAGT	106623
QY	582	HisIleAlaLeuValGlyLysAlaMetSerSerThrGluLysSerTyrGluIleMet	601
DB	106622	CACATAGCGCTGGTTGGCAAGCAATGTCTTCAACCGAGGAGAGTCTCTACGAAATATG	106563
QY	602	ThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly	621
DB	106562	ACATCTCTTGACCTGCATCTGTTTGTGTGATCTTTGGCGGTGTGATCTGCTATTCTGGC	106503
QY	622	AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp	641
DB	106502	GATGATATCAACAGTTCCTGTGGATGTCGCAATTCCTGAGGAGAGCATCCCAAGGAC	106443

QY 642 lleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyVala 661
 Db 106442 ATTAAGGAAGCGATTACTTACCGACCGGCGTGAATTCACGGGTAGATCGCGAAGGTGCT 106383
 QY 662 ProAlaLeuAenCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLys 681
 Db 106382 CCGGCCCTTGCTCACTGCTTATGTACAAATTAAGTACTACAGATTTCGGGGAATTGAAG 106323
 QY 682 LeuAspTyrArg----- 685
 Db 106322 TTGGACTACAG-GTAAGCGGAAACATTTCTCTCAGTAGCGATGCACACTAACTGCCTGT 106264
 QY 686 -----GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
 Db 106263 TTCAGAGTCCATCTGGATATCATCGCACAGTAAACCGGCTCATTTGGGAATAGGACTTC 106204
 QY 704 AspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArg 723
 Db 106203 GATCTGACCTACTCTGGAGGCGCTACACAGAACACTGGCTTGTTCGCATCTATAGG 106144
 QY 724 ValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIlePro 743
 Db 106143 GTGAAGAAGCGCGATGATTCATATAGACCATCACTGAAGACCAAGGAGAGACGATTCCT 106084
 QY 744 ProAlaAsnPheIleSerArgLys----- 751
 Db 106083 CCAGCAACTTCATTCGAGAAAGGTAGTACCTCGACGCTCCCAATGAACCTTCATTAA 106024
 QY 752 -----AsnSerLysArgLysGlyTyrIleArgAsnArgProVal 765
 Db 106023 CAATCTGATCTTTGCAGAACTCTAAGCGTCGCAAGGCTACATACAGAACCGGTT 105964
 QY 766 ValValLysGlyLysArgThrLeuLys 774
 Db 105963 GTTGTTAAGGAAACAGACCTTGAA 105937

RESULT 6

AC008206/c

LOCUS

AC008206 181132 bp DNA linear INV 24-FEB-2001
 Drosophila melanogaster, chromosome 3R, region 96B-96B, BAC clone
 BACR03115, complete sequence.

ACCESSION

AC008206

VERSION

AC008206.10

KEYWORDS

HTG

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 181132)

Celnikier,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,

Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,

Rogers,Y., An,H., Baldwin,D., Banazon,J., Beeson,K.X., Busam,D.A.,

Carlson,J.W., Center,H., Champs,M., Davenport,L.B., Dietz,S.M.,

Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,

Fertler,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,

Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,

Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,

Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,

Phuanavanong,S., Pfitman,G.S., Puri,V., Richards,S., Scheeler,F.,

Stapleton,M., Strong,R., Swirskas,R., Tector,C., Williams,S.M.,

Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

Sequencing of Drosophila chromosome 3R, region 96B-96B

Unpublished

2 (bases 1 to 181132)

Celnikier,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,

Butenhoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,

Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,

Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.

Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,

TITLE

JOURNAL

COMMENT

Swirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
 Rubin,G.M.

Direct Submission

Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Feb 24, 2001 this sequence version replaced gi:7208834.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdgs@fruitfly.berkeley.edu.

FEATURES

source

1..181132

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/strain="Y, cn bw sp"

/db_xref="taxon:7227"

/chromosome="3R"

/map="96B-96B"

/clone="BACR03115 (D765)"

/clone_lib="RPCI-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

ORIGIN

Alignment Scores:

Pred. No.: 1.33e-301 Length: 181132

Score: 3918.50 Matches: 772

Percent Similarity: 90.93% Conservative: 0

Best Local Similarity: 90.93% Mismatches: 2

Query Match: 96.85% Indels: 77

DB: 3 Gaps: 4

US-10-028-384-8 (1-774) x AC008206 (1-181132)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 Db 171813 ATGATCGAGCGCCGAGATGCTGAACAGCAGGTGGCTACACGACCTAATCACC 171754
 QY 21 PheAlaIleLeuLeuIleAlaIleTyrLeuAlaGlyPheSerArgLeuPheAlaValIle 40
 Db 171753 TTCCGCTCCTCTGTAATCGCTGCTGCGCGGATTTTCCTCTCGCCTCTTCGCGCTCATC 171694
 QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
 Db 171693 CGTTTCGAGTCGATATTATCCATGAGTTTTCGCTGTTCACTACCGGCGCACGCGCTAC 171634
 QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTyrPro 80
 Db 171633 ATGCTGCAGATGTTGGTACACTTCTCTCACTGGTTCCAGCGCGCGATGATATCCG 171574
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 171573 CTCGCGCAGGATTGTCGCGGTATCCGCTATCCGCGCTGATGATTACGTCGCGCGGAATC 171514
 QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 171513 CATGGCTGCTGCAGCTACTCAACATACCGGTCCATATTCGTGACATCTCGCTGTTCTGT 171454
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 Db 171453 GCGCGGATCTTCAGTCGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGAGCTGTGG 171394
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 Db 171393 TCCGCGCGCGCGCGCTCTTCGCGCGAGCTTCATCGCCATCGCTGCTGCTACATCAGT 171334
 QY 161 ArgSerValAlaAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180

TITLE

JOURNAL

REFERENCE

AUTHORS


```

171333 AGGTCGGTGGTGGATCGTAAACGAGGGGATTCGATATTCGCTCGAGTTCACC 171274
181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAla 200
171273 TACTTCCTGGTGGGTCGCTCAGTGAAGACTGGATCGGTCTGCTGGCGCGCAGCGCT 171214
201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheLeuLeuLeuLeu 220
171213 TTGTCCTACTCTACATCTGTCGGCTGGCTGGGCTGGCTGGCTGGCTGGCTGGCTGG 171154
221 ProLeuHisValPheValLeuLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSer 240
171153 CCCCTGCACGCTTCCTGCTACTGCTCAATATGGGCGAGTACTCGCGGCTGCTGACACG 171094
241 TyrSerThrPheTyrTrpLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
171093 TACAGCACCTTCTACATCTGGGACTGCTGCTTCTCCATCGAGATCCCTTCGTGGGATTC 171034
261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
171033 CAACCGATACGACCACTGACACATGGCTGGCTGGGAGTGTTCGTCTCTTAAGGCC 170974
281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
170973 GTGGCCACCTTGGCCACTTTGCACTGGTCTGCTCGCGCAACGAGTTCGGAAGCTGTTTC 170914
301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
170913 ATCGTCGGCGGATTCGTGGTGGGCTGGGCTGCTTTGTGGCGCTGCTGGTCTCACCATG 170854
321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
170853 CTGGCGCTTGTGGCGCGCTGGAGTGGAGCGCTTCTACTCGCTGTGGGACTCGCTACGCC 170794
341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
170793 ARGATCCACATTCCTCCATTCATTCGATCGGTGGAGCATTCAGCCACCACTTGGTTCG 170734
361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
170733 TTCTCTTTGATCTGCACATCTGGTGGCTGGCTTCCAGTGGGAGTGGTACTGCATC 170674
381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
170673 AAGCAGATCAACGACGACGCGCTTTCGTGGTCTGTAGCCATCAGTCGGTTCCTACTTC 170614
401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
170613 GCTGGTGTGATGTGGCTTTCGATGTTGACCTCACCGCGGTGGTGTGATGTGGCGGA 170554
421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerLysArgMetGly 440
170553 GTGGCCCTTTCGGACGCTTGGAGTGGTTCCTGCAAGAGATTCGCTTAAGGCAATGGGC 170494
441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThr 460
170493 ACAGCCATAAGCGCAGCAGCCAGTGGATGAAGCTGAGGATTCATTCAGAGAAGACG 170434
461 LeuTyrAspLysAla 465
170433 CTGTACGACAGGT-GAGTTCCTACTAACAACATCCATGGGATTTGATTTTAAATACA 170375
466 -----GlyLysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThr 480
170374 TCGCATTTGAGCGGTGGCAAGCTGAAGCATCGTACTTAAGCATGATGCCAGCAGGATACT 170315
481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMet 500
170314 GGCGTACGCTCCCAACCTCAAGAGATATGTTATTTGGCGCTTCTAATGCTGTGTATGATG 170255
501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerProSerIleValLeu 520

```

RESULT 7

AE003750/c

LOCUS

DEFINITION

complete sequence.

ACCESSION

VERSION

AE003750
Drosophila melanogaster chromosome 3R, section 88 of 118 of the
complete sequence.
AE003750 AE002708 AE014297 AE002644
AE003750.2 GI:23172229

227219 bp DNA linear INV 15-MAR-2004

section 88 of 118 of the

AE003750 AE002708 AE014297 AE002644
AE003750.2 GI:23172229

KEYWORDS

SOURCE
ORGANISM

Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; *Drosophila*.

REFERENCE

AUTHORS

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Burton,R.C., Rogers,Y.H., Blazer,V.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,G.L.,
Abril,J.F., Agbayani,A., An,H.J., Andrews-Frankoch,C., Baldwin,D.,
Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M.,
Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S.,
Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brotter,P.,
Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A.,
Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B.,
Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I.,
Dietz,S.M., Dodson,K., Dunn,P., Durbin,K.J., Downes,M., Dugan-Rocha,S.,
Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,
Ferrieria,S., Fleischmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S.,
Geibart,W.M., Glasser,K., Glodek,A., Gong,P., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Kasper,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Murry,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Stapleton,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodgett,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,M., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

2 (bases 1 to 227219)
Celniker,S.E., Wheeler,D.A., Kronmiller,B., Carlson,J.W.,
Halpern,A., Patel,S., Adams,M., Champe,M., Dugan,S.P., Frise,E.,
Hodgson,A., George,R.A., Hoskins,R.A., Laverly,T., Muzny,D.M.,
Nelson,C.R., Pacle,J.M., Park,S., Pfeiffer,B.D., Richards,S.,
Sodergren,E.J., Svirskaas,R., Tabor,P.E., Wan,K., Stapleton,M.,
Sutton,G.G., Venter,C., Weinstock,G., Scherer,S.E., Myers,E.W.,
Gibbs,R.A. and Rubin,G.M.

Finishing a whole-genome shotgun: release 3 of the *Drosophila*

melanogaster euchromatic genome sequence

Genome Biol. 3 (12), RESEARCH0079 (2002)

22426065

12537568

REFERENCE

AUTHORS

3 (bases 1 to 227219)
Misra,S., Crosby,M.A., Mungall,C.J., Matthews,B.B., Campbell,K.S.,
Hradecky,P., Huang,Y., Kaminker,J.S., Millburn,G.H., Prochnik,S.E.,
Smith,C.D., Tupy,J.L., Whitfield,E.J., Bayraktaroglu,L.,
Berman,B.P., Bettencourt,B.R., Celniker,S.E., de Grey,A.D.,
Drysdales,R.A., Harris,N.L., Richter,J., Russo,S., Schroeder,A.J.,
Shu,Q., Stapleton,M., Yamada,C., Ashburner,M., Gelbart,W.M.,
Rubin,G.M. and Lewis,S.E.

Annotation of the *Drosophila melanogaster* euchromatic genome: a

systematic review

Genome Biol. 3 (12), RESEARCH0083 (2002)

22426069

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

12537572

4 (bases 1 to 227219)

Kaminker,J.S., Bergman,C.M., Kronmiller,B., Carlson,J.,

Svirskaas,R., Patel,S., Frise,E., Wheeler,D.A., Lewis,S.E.,

Rubin,G.M., Ashburner,M. and Celniker,S.E.

The transposable elements of the *Drosophila melanogaster*

euchromatin: a genomics perspective

Genome Biol. 3 (12), RESEARCH0084 (2002)

22426070

12537573

5 (bases 1 to 227219)

Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.

Direct Submission

Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,

Rockville, MD 20850, USA

6 (bases 1 to 227219)

FlyBase

Direct Submission

Submitted (06-SEP-2002) University of California Berkeley, 539 Life

Sciences Addition, Berkeley, CA 94720, USA

7 (bases 1 to 227219)

FlyBase

Direct Submission

Submitted (10-MAR-2004) FlyBase, Harvard University, Biological

Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA

On or before Sep 18, 2002 this sequence version replaced

gi:10729289, gi:7301241

Location/Qualifiers

1. .227219

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/chromosome="3R"

complement(149..6378)

/gene="ssh"

/locus_tag="CG6238"

/map="96B10-96B11"

/db_xref="FLYBASE:FBgn0029157"

complement(join(149..3227,3313..3582,3649..3843,

3907..3972,4265..4656,4722..4967,5864..6378))

/gene="ssh"

/locus_tag="CG6238"

/product="CG6238-RA"

/db_xref="FLYBASE:FBgn0029157"

complement(join(149..3227,3313..3582,3649..3843,

3907..3972,4265..4656,4722..4967,5036..5341))

/gene="ssh"

/locus_tag="CG6238"

/product="CG6238-RB"

/db_xref="FLYBASE:FBgn0029157"

complement(join(875..3227,3313..3582,3649..3843,

3907..3972,4265..4656,4722..4967,5864..5920))

/gene="ssh"

/locus_tag="CG6238"

/codon_start=1

/product="CG6238-PA"

/protein_id="AAF56372.3"

/db_xref="GI:23172230"

/db_xref="FLYBASE:FBgn0029157"

/translation="MALVTQRPSPVAGSCNSDGESEDDGNSKGNDRSECFPAGK

TALVALKDIPPLTQSERRLSTSTSSNSTQNNSDIOHLQSMFYLLQREDTLQNA

VKLESORNSTRVLIVASRCCSGSDRRHRMRHSVKVSGAGSTKSTSPAVT

QRQSVQATPATEASSKCDKADKENATAGDNKNTSGMEESCLLGDICNERTTIGLVV

FILADTTIHLDDGGGFSVKYEKTHIFKPVSVQAMMSALQTLHKVSKARENPFYAGK

PMDWSSYERRIESQSCLENNAMDALSRPPSPDAIRNKPPEKEESVIKMKL

KAIMSVDDLDEVTSKYIRGLIEMDLGCEYKSFDAEMLVILGMDAPTKLFEVY

LGSEWASNLEELQKNGVIRLNVITREIDNFPFGTREVFNVRVYDDEKTNLLKMDT

FRYITRAKESGKLVYCKNKSVSASVITAYAKAYQWEPQALHVKKRSICAPN

KNFLNOLETYSGLVDMKNKELQKRSKSENLKTDARLLPGSEPTPLQAOQSME

KSTGEAGVTDPGEEDGSRHSIAKQSRQRRVWRVRSSTSPKTAFTAVTQKQSQME

NLTPERSVAEPPKMRFPFGSGNGENYSVTQNVLIHQHTPLSVTRTHLEAHRDQL

POQVWTSLLKLTOTSHLGKSVSGSSGNIDSRDSSCSVDVSSQVDSVFAKDBGEX
 RQRKTHSNWTSIESLPGSGVLDPTPOQKQKQSNAILRPRGTQRRLPRRHASWSGDN
 KCKQLKSTSGSYKSNRNTTALFEGVIOQLAKRSNCNVIGVAVAVAFPIVGVGGTV
 KRTQKLEESTSLKRSQKQESQELLEADAGRCPCFSLRYSASAHSPQRQRPLRNC
 SEELHMTSDTENKNTPGDHEATVILRVQGTLDADQRLSGNVQILKQNFRAKHVVG
 TGGGGGTGAAGSGSSSTIATSVAVAGCSANPGKTTSHQSLPSSVPAQKANHVA
 ASNSNSASNSDSSSVSNCNRNVKLVHKYQTTSCSSOAAATPTVACAPLPAGTA
 SNISHSCTTTTTSFKRPLSYDRSLRSHDYSRPPAPRVAPVSHSSNVQHQQQ
 QQOQLOEQHQOQVIRRLAQHCKTHPLSLNAPKTGNTAAYNTM"
 complement(join(875..2227,3313..3582),3649..3843,
 3907..3972,4265..4656,4722..4967,5036..5095))
 /gene="ssh"
 /locus_tag="CG6238"
 /codon_start=1
 /product="CG6238-PB"
 /protein_id="AAAI4027.1"
 /db_xref="GI:23172231"
 /db_xref="FLYBASE:FBgn0029157"
 /translation="MSQVCSLFDLILFSLTPRDESEDDSGNSKGNDRSECEFPAGK
 GTALVALKDIPLTOSRRLSDTSRNSSTQSNNSDIQLHSMFYLLQREDPLKM
 AVKLESQRNRTLYLIASRCCRSQTSRRRHRIMRHSVRVGSAGTKSTSPAPV
 TQRLSVEQTATBASCKDKTADENATAGDNKNTSGMEESCLLIGDCNERTIGLV
 VPIADTTIHLDDGGGSPVKVEKTHIFKPSVOAMWALOTLHKVSKKARENIFYAS
 GPSHDWLSYRESDQSCLENNAMDALSRPPSPDAIRNKPKKEETESVTKMK
 LKAIMSVLDLDEVTKYRCGRLEILDMDLGVKSFIDAEMLVILGQMDAPTKIFHV
 YLGSEWNASNLESLQKGVHILNVTREIDNFPFGTFEYFNVRYVDDKTLNLYKWD
 TFRITKAGAGSKVILVCHGVRSASVVIAYAKAYQWFFQALEHVKRRSCIKP
 NKNFLNGLETYSGLMDAMKXELQKSKSETNLKSTKDARLLPGSEPTPLIQAQNAK
 SKSTGEATVPDGEEDGSMRHSIAQSRMRVRSSTSPKTAVTTKQOSQSM
 ENLTPERSVAEEPKNRFPGSGENYSVTCNOVLHIQKHTPLSVTRIHDLFAHRADQ
 LPQOVVTSKLTITOTSHLGKSVSGSSGNIDSRDSSCSVDVSSQVDSVFAKDEGE
 KRCKRTHTSYSLGSGVLDPTPOQKQKQSNAILRPRGTQRRLPRRHASWSGDN
 NRCLPQRTSSGYSVSNRNTTALFEGVIOQLAKRSNCNVIGVAVAFPIVGVGGTV
 VKEPKLEESTSLKRSQKQESQELLEADAGRCPCFSLRYSASAHSPQRQRPLRC
 NSEBMLSDIENKNTPGDHEATVILRVQGTLDADQRLSGNVQILKQNFRAKHVVG
 GTGGGGGTGAAGSGSSSTIATSVAVAGCSANPGKTTSHQSLPSSVPAQKANHVS
 ASNSNSASNSDSSSVSNCNRNVKLVHKYQTTSCSSOAAATPTVACAPLPAGT
 ASNTSHSCSTTTTTSFKRPLSYDRSLRSHDYSRPPAPRVAPVSHSSNVQHQQQ
 QQOQLOEQHQOQVIRRLAQHCKTHPLSLNAPKTGNTAAYNTM"
 6683..8846
 /locus_tag="CG13645"
 /map="56B11-96B11"
 /db_xref="FLYBASE:FBgn0039254"
 join(6683..6899,6972..7378,7430..7531,7590..7741,
 8330..8600,8709..8846)
 /locus_tag="CG13645"
 /product="CG13645-RA"
 /db_xref="FLYBASE:FBgn0039254"
 join(6683..6899,6972..7378,7430..7531,7590..7741,
 8330..8600,8709..8846)

CDS

Alignment Scores:
 Pred. No.: 1.78e-301 Length: 227219
 Score: 3918.50 Matches: 772
 Percent Similarity: 90.93% Conservative: 0
 Best Local Similarity: 90.93% Mismatches: 7
 Query Match: 96.85% Indels: 22
 Db: 3 Gaps: 4

US-10-028-384-8 (1-774) x AB003750 (1-227219)

Qy 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20

Db 131000 ATGAATCGAGCGCGGAGAGTCTGAAACAGCAAGCGTGGCTGCTACAGCAAGCCTAATCACC 130941

Qy 21 PheAlaIleLeuLeuIleAlaTTPLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40

Db 130940 TTCGCCATCTGCTAATCGCTGCTGCGCCGCGATTTTCTCTCGCCCTCTTCCCGCTGATC 130881

Qy 41 ArgPheGluSerIleIleHisGluPheAspProTTPPheAsnTyrArgAlaThrAlaTyr 60

Db 130880 CGTTTCGAGTCGATATATCCATGAGTTTGATCCGCTGTTCAACTACCGGCGCCACCGCTAC 130821

Qy 61 MetValGlnAsnGlyTTPyrAsnPheLeuAsnTTPPheAspGluArgAlaTTPyrPro 80

130820 ATGCTGCAAGATGGTTGGTACAACTTCTCACTGGTTTCGAGCGAGCGCATGATCCG 130761
 Qy 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 130760 CTCGGCAGGATTTGTGGCGCGTACCGTCTATCCGCGCTGATGATTATCGTCCGGCGGAATC 130701
 Qy 101 HisTTPLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 130700 CATTGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCTCG 130641
 Db 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTTP 140
 130640 GCGCCGATCTTCAGTGGCGCTGACCTCCATCTCCACTACCTGCTGACCAAGAGCTGTGG 130581
 Qy 141 SerAlaGlyValAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 130580 TCCGCGGGCGCGGCTCTTTCGCGCCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT 130521
 Db 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 130520 AGTCTCGTGGCTGGATCGTACGATACGAGGCGCATTCGCATATTCGCCCTGACGTTCCAC 130461
 Qy 181 TyrPheLeuTTPValArgSerValLysThrGlySerValPheTTPSerAlaAlaAla 200
 130460 TACTTCTGTGGTGGCTCAGTGAAGCTGGATCCGTGTTCTGGTCCGCCGCGCGCT 130401
 Db 201 LeuSerTyrPheTyrMetValSerAlaTTPGlyGlyTyrValPheIleIleAsnLeuIle 220
 130400 TTGTCTACTTCTACATGGTGTCCGCTCGGCTGGCTACGTTGTTTCATCATCACTACCTGATA 130341
 Qy 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
 130340 CCCTTCGACGTCTTCGACTGCTCATTTAGGCGAGTACTCCGCGCTGCTGACACGAGC 130281
 Db 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 130280 TACAGCACCTTCTACATCCGCGACTGCTGTTCTCCATGCAGATCCCTCTCGTGGGATTC 130221
 Qy 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 130220 CAACCGATACGACCGAGTGAACACATGCTGGCTGGGAGTGTGTGCTCTTATCGCC 130161
 Db 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 130160 GTGGCCACCTTCGCCCATTTTCAGTCCGTGCTGCGCAACAGAGTTCGCGAGCTGTTTC 130101
 Qy 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValValLeuThrMet 320
 130100 ATGCTCGCGGAGTTCGCTGGTGGCGCTTGGCGCTTTCGCGCGCTGCTGGTGTCCACCATG 130041
 Db 321 LeuGlyValValAlaProTTPSerGlyArgPheTyrSerLeuTTPAspThrGlyTyrAla 340
 130040 CTGGCGGTTGTGGCGCGCTGGAGTGGAGCTTCTACTCGCTGTCGGATCTGGGTACGCC 129981
 Qy 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTTPPheSer 360
 129980 AAGATCCATATCCCATCATTCATCCGTGTCGAGCATCAGCCACCATCTGTTGTTCTCG 129921
 Db 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTTPyrCysIle 380
 129920 TTCTTCTTGTGATCTGCATCTCTGTTGCGCTTCCCGAGTGGAGGTGTGGTACTGATC 129861
 Qy 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 129860 AAGCAGATCAACGACGAGCGGCTTTCGTGGTGTGTGACGCCATCAGTCAGTCGGTTTACTTC 129801
 Db 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 129800 GCTGGTGTGATGCTGCTTGTGATGATCCCTCAGCGCGGTGGTGTGATGCTGCGCGGA 129741
 Qy 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440

129740 GTGGCTTTTCGGACCTGTTGGATGTGTTCTTCGCAAGAGGATTCCTCTTAAGCGAATGGC 129681
 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThr 460
 129680 ACAGCCATAAGCGCAGCCACCGAAGTGGATGAAGCTGAGATTCATTGAGAAGAGACG 129621
 461 LeuTyrAspLysAla 465
 129620 CTGTACGACAGGT-GAGTTCTTACTTAACACACATCCATGGGATTTGATTTAAATACA 129562
 466 -----GlyLysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThr 480
 129561 TCGCATTTGGAGGCTGCAGACCTGAAGCATCTGTAAGCATGATGCCAGCAGGATACT 129502
 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuMetMet 500
 129501 GGCGTCAGCTCCACCTGAGAGATGTTGTTATTTGGCCGTTCTAATGCTGTGATGATG 129442
 501 PheAlaValHisCysThrTyrValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
 129441 TTCGCTGTCACCTGACGCTGGTGACAGCAATGCTACTCCAGTCCCTCCATTGTCCTG 129382
 521 AlaPheHisAsnSerGlnAspGly----- 528
 129381 GCTTTCACACACAGTCAAGATGGTGAAGCACCAGTAGCATCATGCTCTTTAAACCTA 129322
 529 -----SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyr 541
 129321 TCTAACCTTTTCTCCACTAGATCCCGCAATTTTACAGCATTTTCAGAGCGCTTACTAC 129262
 542 TrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTyrAspTyrGlyTyr 561
 129261 TGGCTTTCGACAGACACTGCGGATGATGCTCGGTTATGCTTGGTGGGATTCGGATAC 129202
 562 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 581
 129201 CAGATAGCGGAATGGCAACAGACAGCGTAGTGGATTAATAACGTGGAGAACAAATAGT 129142
 582 HisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIleMet 601
 129141 CACATAGCGCTGGTTGGCAAGGCAATGCTCTTCAACCGAGAGAGTCTTACGAAATATG 129082
 602 ThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 621
 129081 ACATCTCTGACGGGACTACGTTTGGTGATCTTTGGCGGGTGATCGGCTATTCTGGC 129022
 622 AspAspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAsp 641
 129021 GATGATATCAACAAGTTCTCTGTGATGGTCCGAATTTGCTGAGGAGAGCATCCCAAGGAC 128962
 642 IleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluAla 661
 128961 ATTAAAGAAACGGATTAATTTACGACCGCGGTAATTCAGGTAGATCGCGAAGTGCT 128902
 662 ProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLys 681
 128901 CGGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCCGGGAATTGAAG 128842
 682 LeuAspTyrArg----- 685
 128841 TTGGACTACAG-GTAAGCGGMAACATTTCTCAGGTAGCGATGCACATACTCCCTGT 128783
 686 -----GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
 128782 TTCAGAGGTCCATCTGGATATGATCGCACACGTAAACCGCTCATTTGGGAATAAGGACTTC 128723
 704 AspLeuThrTyrLeuGluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArg 723
 128722 GATCTGACCTACCTGGAGAGGCCCTACACACAGAACACTTGGCTTTCGCACTATTAGG 128663
 724 ValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIlePro 743
 128662 GTGAAGAACCGCATAGGTTCAATAGACATCACTGAACCAAGCAGAGAACGATTCCT 128603
 743 ValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIlePro 743
 128603 GTGAAGAACCGCATAGGTTCAATAGACATCACTGAACCAAGCAGAGAACGATTCCT 128603

Db 19030 GAATCTGATCTTTTCAGAACCT-AAGCGTCCAGGCGCTACATACGAAACCGCGCTT 18972

QY 766 ValValysGlyLeuArgThrLeuLys 774

Db 18971 GTTGTAAAGGAAACGACCTTGAA 18945

RESULT 9

LOCUS BC052433

DEFINITION Mus musculus RIKEN cDNA 1300006C19 gene, mRNA (cDNA clone MGC:64679 IMAGE:6837097), complete cds.

ACCESSION BC052433

VERSION BC052433.1 GI:30851501

KEYWORDS MGC

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 4236)

AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.D., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, J.A., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 2388257

PUBMED 12477332

REFERENCE 2 (bases 1 to 4236)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (15-MAY-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr. Thomas L. Casavant.
Web site: <http://genome.uiowa.edu>
Contact: bento-soaresuiowa.edu; tom-casavant@uiowa.edu
Bonaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A., Fisher, K., Kappel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K., Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J., Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.
Location/Qualifiers
1. .4236
/organism="Mus musculus"

FEATURES
source

gene

CDS

misc_feature

ORIGIN

Alignment Scores:

Pred. No.: 1,84e-226 Length: 4236

Score: 2954.00 Matches: 552

Percent Similarity: 83.27% Conservative: 90

Best Local Similarity: 71.60% Mismatches: 113

Query Match: 73.01% Indels: 16

DB: 10 Gaps: 5

US-10-028-384-8 (1-774) x BC052433 (1-4236)

QY 10 SerLysValAlaGlyTyrSerSerLeuIleThrPheAlaIleLeuLeuIleAlaTyrLeu 29

Db 543 TCGACGCGCGCGCGTGGCAGTCGTCCTCTCCCTCCACATCCCTCTTCCCTGGCTGCTG 602

QY 30 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHisGluPhe 49

Db 603 GCCCGCTTCAGTCGCGCTCTTCGCGCTTCATCGCTTCGAGAGCATCATCCAGATTC 662

QY 50 AspProTyrPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPhe 69

Db 663 GACCGGTGGTTAACTATAGATCAACATCATCTTCATCTCATGGATTTCTTAGATTT 722

QY 70 LeuAsnTyrPheAspGluArgAlaTyrTyrProLeuGlyArgIleValGlyIleThrVal 89

Db 723 CTAAATTTGGTTTGTATGAAGAGCATGTTACCCCTCGGAAGATAGTGGTGGCCCGTT 782

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="MGC:64679 IMAGE:6837097"

/tissue_type="Brain, enriched mouse brain 12.5dpc"

/clone_lib="NIH_BMAP_F10"

/lab_host="DH10B"

/note="Vector: pYX-ASC"

1. .4236

/gene="1300006C19Rik"

/notes="synonym: Simp"

/db_xref="LocustID:68292"

/db_xref="MGI:1915542"

366. .2837

/gene="1300006C19Rik"

/cdon_start=1

/product="1300006C19Rik protein"

/protein_id="AAH52433.1"

/db_xref="GI:30851502"

/db_xref="LocustID:68292"

/db_xref="MGI:1915542"

/translation="MAEPADESKHKSINSPPSLMALGNSRHGHGPGTQSASRA
AAPKPGPAGLSQAGWQSLFLLFLAWAGFSSRLFAVRESIHEDPWP
FYRSTHLLASHGFYEFLLNFDERAWYPLRGVGTVPGLMTAGLHWLINTLIT
VHIRDVCVLAFTFSGLTSTIFLLTRLMQAGLAACTAIYVGISSVAGSFD
NEGIAIFALQFTYLYWKSIVTGSVFTWCCLSYFVWSAWGGVFIINILPLAFV
LLLMORYSKRVYIAYSTFIVGLILSMOIPFVGPOPIRTSHMAAGVFAILOAVFL
QYLRDLTKQRFOTLFTGLVSLAAGAVLSVLYTYTGVIAPWSGRFYSLNDTGAKI
HIPITASVSEHQPTVWVSFAFNSVFEHGLDDMKRENPPVEDSDDEKRNPNGLY
AGWVRMLTIFVCMVSAIAFNSVFEHGLDDMKRENPPVEDSDDEKRNPNGLY
DAGKVRKHVTEQEPGLGNPKTSIVTLMMLMMFAVCTWVTNAYSVPGLV
ASYNHGTENLDDFREYFWLRQNTDEHARVMSWDYQYIAGMANTTLVDNNTW
NSHALVGVKSSNETAAYKMRSLDVLVIFGVIGYSGDDINKKTLWVRJAEGE
HPKIDREGDYFTQCGEPVVDKAGSLTLANCYKMSYVFGEMOLDPEPTPGFDTRN
AEIGNKDIKFKHLEAFSEHMLVRIYKAPDRETLCHKPRVTNIVPKKILSKKT
TKRKGYYVKNLKVFKGKSKTKIV"

567. .2537

/gene="1300006C19Rik"

/note="STT3; Region: Oligosaccharyl transferase STT3
subunit. This family consists of the oligosaccharyl
transferase STT3 subunit and related proteins. The STT3
subunit is part of the oligosaccharyl transferase (otase)
complex of proteins and is required for its activity.
Otase transfers a lipid-linked core-oligosaccharide to
selected asparagine residues in the ER"

/db_xref="CDD:pfam02516"

Qy 90 TyrProGlyLeuMetIleThrSerGlyGlyIleHisTripleLeuHisValLeuAsnIle 109
 Db 783 TACCAGGTTGATGATAACAGCTGGCTTATTCATTGGATTAAATACATTGAACATA 842
 Qy 110 ProValHisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer 129
 Db 843 ACAGTTCATCAATAGAGATGTGTGTATTCCTTTCACCAACTTTTACGGCGCTTACATCC 902
 Qy 130 IleSerThrTyrLeuLeuThrLysGluLeuTyrSerAlaGlyValAlaGlyLeuPheAla 149
 Db 903 ATATCTAGTTCTCTGCTAATAGAGAACTGTGGAAACCAAGGACAGACTTCTAGCTGCT 962
 Qy 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 169
 Db 963 TGCCTTCATTGTATCGTACAGAGGTACATATCTCGTCACTGGCGGATCCTTTGATAAT 1022
 Qy 170 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTyrValArgSerValLys 189
 Db 1023 GAAGGATTCGCAATTTTGGCTTCAGTTTCACTTACTTATGGGTAAAGTCTGTGAAG 1082
 Qy 190 ThrGlySerValPheThrSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAla 209
 Db 1083 ACCGGTCTGTGTTCTGGCAAAATGTGCTGTCTGTCTCATATTTTACATGTCTCTGCG 1142
 Qy 210 TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIle 229
 Db 1143 TGGGAGGTATGTTTTCATCATCACTCATCCCTTCCATGTGTGTGTGTGTGTGTGT 1202
 Qy 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerTyrPheTyrIleLeuGlyLeu 249
 Db 1203 ATGCAGAGGTACAGCAAGAGAGTCTACATAGCATATAGCATTTTCTACATTTGTGGTTA 1262
 Qy 250 LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
 Db 1263 ATATTATCCATGCAGATACCTTTTGTGGATTTTACGCCATCAGAACAGCGAGCATG 1322
 Qy 270 AlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSer 289
 Db 1323 GCACCTGCAGGTGTCTTTTGGCTGTCTGCAAGCTTACGCTTTTGTGAGTATCTGAGAC 1382
 Qy 290 ValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyLeuLeuValGlyVal 309
 Db 1383 CGGTTGACAAACAGAGATTTCCAGACCTTTTCTTTTGGGTCTCTACCTAGCTCCAGC 1442
 Qy 310 GlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProTyrSerGly 329
 Db 1443 GCTGTGTCTTGTATGTCTATCTCTGACATACACAGGTATATTCACCATGGAGTGGC 1502
 Qy 330 ArgPheTyrSerLeuTyrAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSer 349
 Db 1503 AGGTTTATTCATCTAGGATCTGGTATGCAAAATATACATCCATTTATTCATCA 1562
 Qy 350 ValSerGluHisGlnProThrTyrPheSerPhePheAspLeuHisIleLeuVal 369
 Db 1563 GTGCTGAACATCAACGCTACGATGGGTGTCTTTCTTTTGTATCATATTTCTGTA 1622
 Qy 370 CysAlaPheProValGlyValTyrCysIleLysGlnIleAsnAspGluArgValPhe 389
 Db 1623 TGTACCTTCCAGCAGAGCTTATGGTCTGCATCAAAATATCAACGATGAAGAGATATT 1682
 Qy 390 ValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 409
 Db 1683 GTCCTCTGTATGGATTCAGTGTGTCTTTTGGCGGAGTGTGGTGTGGTGTGTGTGT 1742
 Qy 410 ThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspVal 429
 Db 1743 ACTCTGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1802
 Qy 430 PheLeuGlnGluAspSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluVal 449
 Db 1803 TATTGGGGATGATGATGAAGAGGAA-----AACCCACCTGTG 1841

Qy 450 AspGluAlaGluAspSerIleGluLysLys-----ThrLeuTyrAspLysAlaGly 466
 Db 1842 GAGGACAGCAGTGTAGGATGACAAAGAAACCCAGGAAACTTTGTATGACAGGAGGT 1901
 Qy 467 LysLeuLysHisArgThrLysHisAspAlaGlnAspThrGlyValSerSerAsnLeu 486
 Db 1902 AAAGTGAAGAAGCATGTGACAGCAAGAAACCTGAAGAGGCTTGGGCCCCAATC 1961
 Qy 487 LysSerIleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThr 506
 Db 1962 AAAAGCATGTGACCATGCTGATGCTCATGCTCTCTGTATGATGTTCGGTCCCATGCAAG 2021
 Qy 507 TrpValThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGln 526
 Db 2022 TGGGTCAACAGCAACGCTTACTCCAGTCCAAAGTGGTCTTGGCTTCTCTCAAT--CAT 2078
 Qy 527 AspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsn 546
 Db 2079 GATGTGTACAGCAATATATAGATGATTTTGAAGAAGCTACTTTTGGCTGAGACAAAC 2138
 Qy 547 ThrAlaAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMet 566
 Db 2139 ACGATGAACACGCGCGGCTCATGTGCTGGTGGAGTACGCTATCAGATTGCTGGCATG 2198
 Qy 567 AlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuVal 586
 Db 2199 GCCAACAGGACCATCTGTGTGATTAACACACCTGGAAACACACGACCATGCTGCTGTC 2258
 Qy 587 GlyLysAlaMetSerSerThrGluLysSerTyrGluIleMetThrSerLeuAspVal 606
 Db 2259 GGAAGAGCTATGTCTTCCAAAGAACCGCGCCCTATAAATCATGAGTCCCTTGATGTC 2318
 Qy 607 AspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLys 626
 Db 2319 GATTATGTGTGTATTTTCGAGAGTGTATGCTGCTATTCGCGGAGCATATCAACAAG 2378
 Qy 627 PheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAsp 646
 Db 2379 TTCTCTGTGATGGTCAGATAGCTGAGGGAGGAGCATCCAAAGACATCCGGGAGGTGAC 2438
 Qy 647 TyrPheThrAspArgGlyGlyPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsn 666
 Db 2439 TATTTTCCACGACGAGGAGAGTTCGAGTAGACAAAGCTGGTCTCTCTACTCTGTTAAAC 2498
 Qy 667 CysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGly 686
 Db 2499 TGCCTTATGATAAAATGTCTACTACAGATTTGGAGAAATGCGAGCTAGATTTTGCAC 2558
 Qy 687 ProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThr 706
 Db 2559 CCCCCAGGCTTTGACCGAACACAGTAATGCTGAGATTGGAATAAGACATTAAATTCAG 2618
 Qy 707 TyrLeuGluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLys 726
 Db 2619 CATTTGGAGAAAGCTTTTACATCAGACGACCTGGCTTGTGAGGATATATAAGTGAAGCA 2678
 Qy 727 ProHisGluPheAsnArgProSerLeuLysThrLysGluArgThr-----IlePro 743
 Db 2679 CCTGAC-----AACAGGAGACACTAGTCCAAACCTCGAGTCACCAACATGCTCCCC 2732
 Qy 744 ProAlaAsnPheIleSerArgLysAsnSerLysArgLysGlyTyrIleArgAsnArg 763
 Db 2733 AACACGAAGATTTTCTCAAGAAAGACTACTAAAGAGAGCGTGGCTACGTTAAAAAATAG 2792
 Qy 764 ProValValValLysGlyLysArgThrLeuLys 774
 Db 2793 CTAGTGTTTAAGAAAGGCAAGAGACCTCTAAG 2825

RESULT 10
 AX799082
 LOCUS
 DEFINITION
 Sequence 1 from Patent WO0305408.
 AX799082
 ACCESSION

AX799082 2481 bp mRNA linear PAT 08-OCT-2003

QY	10	SerLysValAlaGlyTyrSerSerLeuIleThrPheAlaIleLeuLeuIleAlaTirPLeu	29
DB	187	TGCGACCGCGTGGGTGGCAGTCGCTTCTCCCTCCACCATCTCTCTGGCGCTGGCAT	246
QY	30	AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe	49
DB	247	GCGGCTTCAGCTCCGCGCTCTTCGCGCTCATCGCTTCGAAGCATCATCCAGATTC	306
QY	50	AspProTirPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTirTyrAsnPhe	69
DB	307	GACCGCTGGTTTAACTATAGATCAACACATCATCTGCATCTCATGGGTTCTATGAATTT	366
QY	70	LeuAsnTirPheAspGluArgAlaTirTyrProLeuGlyArgIleValGlyGlyThrVal	89
DB	367	TTAAATTTGGTTTGATGAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGTACTGTT	426
QY	90	TyrProGlyLeuMetIleThrSerGlyGlyIleHisTirPLeuIleHisValLeuAsnIle	109
DB	427	TACCGAGGGTGGATGAATACCGCTGGCTTATTCATTGGATTTTAAATACATTGAACATA	486
QY	110	ProValHisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer	129
DB	487	ACTGTTTCACATAAGAGAGCTATGTGTCTCTTCGACCACTTTTAGCGGCTTACATCT	546


```

Db 1612 ATGTGACCATGTTGATGCTGATGCTATGATGATGTTGCTGCTCCACTGATCCTGGGTC 1671
Qy ThrSerAsnAlaTyrSerSerProSerIleValleuAlaPheHisSerGlnAspGly 528
Db 1672 ACAAGCAATGCTACTCTAGTCCAAAGGTAGTCTCGCCCTCATACAAT--CATGATGCG 1728
Qy 529 SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAla 548
Db 1729 ACCAGCAATATCTTAGATGATTTTAGAGAGCTTACTTTTGGCTTAAGGCAAAATACAGAT 1788
Qy 549 AspAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 568
Db 1789 GAACATGACAGGAGTAATCTCTGTGGGATATGCTATCAGATAGCTGGAATGGCTAAT 1848
Qy 569 ArgTrpThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 588
Db 1849 AGAATCTAGCTTGGTGATATAACACCTGGAATAACAGCCACATAGACATGTTGGGAAA 1908
Qy 589 AlaMetSerSerThrGluGlySerTyrGluIleMetThrSerLeuAspValAspTyr 608
Db 1909 GCTATGCTCTTAATGAACAGCAGCCTATAAATCATGAGGACTCTAGATGTAGATTAAT 1968
Qy 609 ValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 628
Db 1969 GTTTTGGTATTTTGGAGGGTATTGGCTATTCTGTGATGATATCAACAAATTTCTC 2028
Qy 629 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
Db 2029 TGGATGGTTAGATAGCTGAGAGGAGACATCCCAAGACATTCGGGAAATGACTATTTT 2088
Qy 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
Db 2089 ACCCCACAGGAGGAAATTCGGTGTAGACAAACAGGATCCCTACTTTGTTGATTCCTT 2148
Qy 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
Db 2149 ATGTATAAATGTATACATCAGATTTGGAGAAATGACGTGGATTTTGGTACACCCCA 2208
Qy 689 GlyTyrAspArgTrpArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
Db 2209 GGTTTTGGCCGACACAGTAATGCTGAGATTTGGAATAAGGACATTAATTCAAACATTG 2268
Qy 709 GluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHis 728
Db 2269 GAAGAAGCCTTTACATCAGAACACTGGCTGTAGGATATATAAGTAAAGACCTGAT 2328
Qy 729 GluPheAsnArgProSerLeuLysThrLysGluArgThr-----IleProProAla 745
Db 2329 -----AACAGGAGACATTAGATCACAACCTCGAGTCACCAACATTTTCCCAAAACAG 2382
Qy 746 AsnPheIleSerArgLysAsnSerIysArgArgLysGlyTyrIleArgAsnArgProVal 765
Db 2383 AAGTATTTGTCAAGAGAGCTACCAAGGAGGCGTGGCTACATTAATAAATAGCTGGTT 2442
Qy 766 ValValLysGlyLysArgThrLeuLys 774
Db 2443 TTTAAGAAAGGCAAGAAATATCTAAG 2469

```

RESULT 11

```

AY074880
LOCUS 2481 bp mRNA linear PRI 10-DEC-2002
DEFINITION Homo sapiens source of immunodominant MHC-associated peptides
(SIMP) mRNA, complete cds.
ACCESSION AY074880
VERSION AY074880.1 GI:19879588
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2481)
AUTHORS McBride,K., Baron,C., Picard,S., Martin,S., Boismenu,D., Bell,A.,

```

Bergeron,J. and Perreault,C.
 The model beta1 minor histocompatibility antigen is encoded by a
 mouse homolog of the yeast STM3 gene
 Immunogenetics 54 (8), 562-569 (2002)
 22326278
 PUBMED 12439619
 REFERENCE 2 (bases 1 to 2481)
 AUTHORS McBride,K. and Perreault,C.
 TITLE Direct Submission
 JOURNAL Submitted (22-JAN-2002) Molecular Biology Group, Compatisgene, 6100
 Royalmount, Montreal, Qc H4p 2R2, Canada

FEATURES

source
 1..2481
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 1..2481
 /gene="SIMP"
 1..2481
 /genes="SIMP"
 /note="similar to Saccharomyces cerevisiae OSTTT3 and Mus
 musculus protein encoded by GenBank Accession Number
 AK018758; source of minor histocompatibility antigen(s)"
 /codon_start=1
 /product="source of immunodominant MHC-associated
 peptides"
 /protein_id="AA171884.1"
 /db_xref="GI:19879589"
 /translation="MAEPSPAPAGSLGSLQSPWOSLSFTILFLAWAGFSSRLFAVIFPESITHEF
 AGAPAPKPAAGSLGSLQSPWOSLSFTILFLAWAGFSSRLFAVIFPESITHEF
 DFWNYRSTHLSHGFEFLNDFEDRWYPLRGIVGGTVVPGMLTGLHMLNTL
 NITVHRDVCVFLAFSGLTSTISFTILRELNNQAGLLAACFIALVPGYISRSVAG
 SPNEGIALFALOFTYVLWKSVKTSVLMCCCLSVFYVWSAWGVYVILNLIPLH
 VFVLLMORYSKRVYIAYSTFYIYGLIMOIPEVGPQIRTSHEHMAAGVALLQAY
 APLVRLDLTKQBEFTLFLGVSLSAAGAVELSVLYLTGYIAPNSGRFVSVDWGY
 AKHLPITLASVEHGFPTTWSFFPDLHLVCTPAGLWFCIKINIDRERVVALYISA
 VFAGVYVNMULTLPVVMSLATAFNSVFEHGLGDDMKRENPPVSDSDDDRNGG
 NYDKAGVGRGHATQEETEGELGNIKSVITMLMLLMFMFAVITWVTSNAYSSPS
 VLYASVNDHTNILLDFREAYFWLRQNTDEHARVNGWDYGYIAGMANRTLLVNN
 TWNNHIALVKGMSNETAAYKIMRLDVLVIVFGVIGSGDDINKFLMWRIA
 EGEHPADRESYFTTPOGEFRVDKAGSPITLNCILMKYMSYVRFGEMOLDPTPPGPR
 TNAETGNCKDIKFHELEAFTSEHMLVRIYKVKAPDNRETLDHKPRVTNIFPKQKYL
 KKTTEKRGYIKKLVFKKGIKSKTV"

Alignment Scores:
 Pred. No.: 2,32e-226 Length: 2481
 Score: 2949.00 Matches: 556
 Percent Similarity: 83.88% Conservative: 89
 Best Local Similarity: 72.30% Mismatches: 112
 Query Match: 72.89% Indels: 12
 DB: 9 Gaps: 7

ORIGIN

US-10-028-384-8 (1-774) x AY074880 (1-2481)
 Qy 10 SerLysValAlaGlyTyrSerSerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeu 29
 Db 187 TCGCAGCGCGTGGTGGCAGTCGCTTCTCTCTCCATCCATCTCTCTCGCTGCGCTT 246
 Qy 30 AlaGlyPheSerSerArgLeuPheAlaValIleAlaArgPheGluSerIleIleHisGluPhe 49
 Db 247 GCCCGGTTTCAGCTCGCGCTCTTCGCCGCTCATCCCTCCGAAAGCATCATCCAGAGTTC 306
 Qy 50 AspProTrpPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTyrTrpTyrAsnPhe 69
 Db 307 GACCCGCTGTTTAACATATAGTCAACACATCATCTTGCATCTCATGGTCTTCAATGATTT 366
 Qy 70 LeuAsnTrpPheAsnGluArgAlaTrpTyrProLeuGlyArgIleValGlyIleThrVal 89
 Db 367 TTAATATGGTTTGAATAAGAGCATGGTATCCATAGGAAGATAGTAGGTGGTACTGTT 426
 Qy 90 TyrProGlyLeuMetIleThrSerGlyIleHisTrpLeuLeuHisValLeuAsnIle 109

Db 427 TACCAGGCTTGCATACCGCTGGCCTTATTCATTGGATTTTAAATACATTGAACATA 486
Qy 110 ProValHisIleAetAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer 129
Db 487 ACTGTTTCATACAGAGCTATGTGTCTTCACCACTTTAGCGCCCTTACATCT 546
Qy 130 IleSerThrTyrLeuLeuThrLysGluLeuThrPheAlaGlyAlaGlyLeuPheAlaAla 149
Db 547 ATATCTACTTTCCTGCTTACAGAGAACTTTGGAAACCAAGAGCAGGACTTTTAGCTGCT 606
Qy 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 169
Db 607 TGTATTTATGCTATTTGACAGCTACATATCTGGTCAGTAGCTGGATCTTTGATPAT 666
Qy 170 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuThrPheAlaArgSerValLys 189
Db 667 GAAGGCATTCGCTATTTTGCACCTTCAGTTCACATACTATTTATGGGTAAATCTGTAAAA 726
Qy 190 ThrGlySerValPheThrPheAlaAlaAlaLeuSerTyrPheTyrMetValSerAla 209
Db 727 ACTGGGTCAATTTTGGACATGTGCTGCTTAICCTATTTCTATATGGTCTCTGCT 786
Qy 210 TrpGlyGlyTyrValPheIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 229
Db 787 TGGGTGGTGTATGTATTTATCATCAATCTTATTCACCTGCATGATTTATTTGTTGTTACTG 846
Qy 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeu 249
Db 847 ATGCAGATACAGCAAGAGCTCTACATACATATACATCTTCTACATTTGCGGTGTTA 906
Qy 250 LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
Db 907 ATATTATCAATGCATACATCTTTTGGGATTCACCAATTCAGCAACAAAGTGAACATG 966
Qy 270 AlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSer 289
Db 967 GCAGCTGCGAGGTGCTTTGCTATGCTGCAAGCTTATGCTTCTTGGAGTATCTGAGAGAC 1026
Qy 290 ValLeuSerArgAsnGluPheArgLysLeuPheLeuValGlyLeuLeuValGlyVal 309
Db 1027 CGATTAAACAAACAGAGTTCAGACCCCTTTCTTTTGGGTGTATCACTAGCTGCAGGT 1086
Qy 310 GlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGly 329
Db 1087 GCTGTGCTTCTTATGCTATCTTATTTGACTTATACAGGTTACATTTGACCATGAGTGGC 1146
Qy 330 ArgPheTyrSerLeuThrPheThrGlyTyrAlaLysIleHisIleProIleAlaSer 349
Db 1147 AGGTTTTTATCTTGGGATCTGGGTATGCAAAATATACAAATTTCAATTTATGTCATCA 1206
Qy 350 ValSerGluHisGlnProThrThrPheSerPhePhePhePhePhePhePhePhePhe 369
Db 1207 GTGCTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTGTATCTACATTTCTTGTA 1266
Qy 370 CysAlaPheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPhe 389
Db 1267 TGTACCTTCCAGCGCCCTTTGGTCTGCTCAAAATAATCAACGATGAAGAGTATTT 1326
Qy 390 ValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 409
Db 1327 GTTGCTATATGCAATAGTCTGCTACTTTGCTGAGTGTGGTGGAGTGTGGTGGAGTGTG 1386
Qy 410 ThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspVal 429
Db 1387 ACTTTGACTCCAGTCGTGTATGCTGTGCAATTTGCTTTTCAATGTTTTTGGACAC 1446
Qy 430 PheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluVal 449
Db 1447 TATTTCGGGGATGAC---ATGAAAGGGGAAAAATCCACTGCTGGAGCAGCAGGATGAG 1503
Qy 450 AspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeu--- 468

Db 1504 GAT-----GACAAAAGAAACCAAGGAAATTTGTATGATAAGGACGTAAGTGAGG 1554
Qy 469 LysHisArgThrLysHisAspAlaGlnAspThrGlyValSerSerAsnLeuLysSer 488
Db 1555 AAAACATCAACTGACAGGAAAAAAGTGAAGAG---GGATTAGCCCTTAATAAAGAGC 1611
Qy 489 IleValIleLeuAlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpVal 508
Db 1612 ATTTGTCACCATGTTGATGCTGATGCTTATGATGCTTTCTGCTGCTGCTGCTGCTGCTG 1671
Qy 509 ThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly 528
Db 1672 ACAGCATGCTTACTCTAGTCCCAAGGTAGTCTGCTGCTCATACAAT---CATGATGC 1728
Qy 529 SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAla 548
Db 1729 ACCAGGAATATCTTAGATGATTTTGGAGAGCTTACTTTTGGGTAAAGGCAAAATACAGAT 1788
Qy 549 AspAspAlaArgValMetSerTrpTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 568
Db 1789 GAACATGACCCAGTATGCTTGTGGGATTTAGCTATCAGATAGCTGGAATGGCTAAT 1848
Qy 569 ArgThrThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 588
Db 1849 AGAATACCTGCTGGTGAATAACACCTGGAATAACAGCCACACATAGCAGCTGGTGGGAAA 1908
Qy 589 AlaMetSerSerThrGluLysSerTyrGluIleMetThrSerLeuAspValAspTyr 608
Db 1909 GCTATGCTTCTTAATGAACAGCAGCCTATAAATCATGAGGACTCTAGATGTAGATTAT 1968
Qy 609 ValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 628
Db 1969 GTTTTGGGTATTTTGGAGGGGTTATTGGCTATTCTGCTGATGATATCAACAAATTTCTC 2028
Qy 629 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
Db 2029 TGGATGTTAGATAGCTGAGAGAGACATCCCAAGACATTCGGGAAGTACTATTTT 2088
Qy 649 ThrAspArgGlyGluPheArgValAspAlaGluAlaProAlaLeuLeuAsnCysLeu 668
Db 2089 ACCCCACAGGGAGAAATCCGTGTAGACAAAGCAGGATCCCTCTACTTTTGTGAATGCTT 2148
Qy 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
Db 2149 ATGTAATAAATGTCTACTACAGATTTGGAGAAATGAGCTGGATTTTCTACACCCCA 2208
Qy 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
Db 2209 GGTTTTGACCCGACACAGTAAATGCTGAGATTTGGAATAAGGACATTAATTTCAACATTTG 2268
Qy 709 GluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHis 728
Db 2269 GAAGAGAGCTTTACATCAGAACACTGGCTTTGAGGATATATAAAGTAAAGACCTGAT 2328
Qy 729 GluPheAsnArgProSerLeuLysThrLysGluArgThr-----IleProProAla 745
Db 2329 -----AAACAGGAGACATTTAGATCAGAAACCTCGAGTCCACACATTTTCCCAAAACAG 2382
Qy 746 AsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765
Db 2383 AAGTATTTGTCAAGAGACTCCCAAAAGGAGGCTGGCTACATTAATAAATAGCTGTT 2442
Qy 766 ValValLysGlyLysArgThrLeuLys 774
Db 2443 TTTAAGAAAGCAAGAAATATCTAAG 2469

RESULT 12

AX799084 LOCUS 2710 bp mRNA linear PAT 08-OCT-2003
DEFINITION Sequence 3 from Patent WO03054008.
ACCESSION AX799084
VERSION AX799084.1 GI:37605059
KEYWORDS

Qy 607 AspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLys 626
 Db 2030 GATTATGTTGGTTATTTTCGGAGGAGTGTATTCGCTATTCGCGGGACGATATCAACAAG 2089
 Qy 627 PheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAsp 646
 Db 2090 TTCCTCTGGATGTCAGATAGCTGAGGGGAGCATCCCAAGACATCCGGGAGGTGAC 2149
 Qy 647 TyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsn 666
 Db 2150 TATTTCACCCAGCAGGAGGAGTTCGAGTAGACAAAGCTGGTCTCTACTCTGTTAAAC 2209
 Qy 667 CysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGly 686
 Db 2210 TGCCTTATGATAAATGTCATCTACAGATTTCGAGAAATGACGTAGATTTCGACT 2269
 Qy 687 ProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThr 706
 Db 2270 CCCCAGGCTTGAACCAACACGTAATGCTGAGATTGGAATAAAGACATTAATTAAC 2329
 Qy 707 TyrLeuGluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLys 726
 Db 2330 CATTTGAGGAAGCTTTTACATCAGACACTGGCTTCAGGATATATAAGTGAAGCA 2389
 Qy 727 ProHisGluPheAsnArgProSerLeuLysThrLysGluArgThr-----IlePro 743
 Db 2390 CCTGAC-----AACAGGGAGACACTAGGTCACAAACCTCGAGTCACCAACATCGTCCC 2443
 Qy 744 ProAlaAsnPhelSerArgLysAsnSerLysArgLysGlyTyrIleArgAsnArg 763
 Db 2444 AACAGAAAGTATTTGTCAAGAGACACTATAAAGGAAGCGTGGCTACGTTAAAAATAAG 2503
 Qy 764 ProValValValLysGlyLysArgThrLeuLys 774
 Db 2504 CTAGTGTTTAAGAAAGCAAGAGACCTCTAAG 2536
 RESULT 13
 UI3019
 LOCUS 54118 bp DNA linear INV 13-JAN-2003
 DEFINITION Caenorhabditis elegans cosmid T12A2, complete sequence.
 ACCESSION UI3019
 VERSION UI3019.1 GI:912482
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 54118)
 Waterston,R.
 Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 MEDLINE 99069613
 PUBMED 9851916
 REFERENCE 2 (bases 1 to 54118)
 Latreille,P.
 The sequence of C. elegans cosmid T12A2
 Unpublished (2001)
 REFERENCE 3 (bases 1 to 54118)
 Waterston,R.
 Direct Submission
 Submitted (06-MAR-1996) Genome Sequencing Center, Washington
 University
 REFERENCE 4 (bases 1 to 54118)
 Waterston,R.
 Direct Submission
 Submitted (26-JUN-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 5 (bases 1 to 54118)
 Waterston,R.
 Direct Submission
 Submitted (09-AUG-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 6 (bases 1 to 54118)
 Waterston,R.
 Direct Submission
 Submitted (19-APR-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 7 (bases 1 to 54118)
 Waterston,R.
 Direct Submission
 Submitted (28-AUG-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 8 (bases 1 to 54118)
 Waterston,R.
 Direct Submission
 Submitted (21-NOV-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 9 (bases 1 to 54118)
 Waterston,R.
 Direct Submission
 Submitted (13-JAN-2003) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 COMMENT
 On Jul 27, 1995 this sequence version replaced gi:529354.
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=T12A2;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C23G10, 200 bp overlap; the 3' cosmid is C18F10,
 15000 bp overlap. Actual start of this cosmid is at base position 1
 of T12A2; actual end is at 39122 of T12A2.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 Genefinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yuji Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
 elegans ORFeome cloning project (<http://wormdb.dfci.harvard.edu/>),
 similarity to other proteins from BlastX analyses
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000), individual C. elegans GenBank submissions,
 and personal communications with C. elegans researchers. tRNAs
 are predicted using the program tRNAscan-SE (Lowe, T.M. and
 Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
 Location/Qualifiers

FEATURES

```

1. 54118
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N239"
/db_xref="taxon:6239"
/chromosome="III"
/clone="T12A2"
3241..6926
/gene="hot-4"
/join(3241..3365,3586..3749,6392..6561,6795..6926)
/gene="hot-4"
/standard_name="T12A2.16b"
/notes="coded for by the following C. elegans cDNAs:
yk1067e11.3, yk1325d06.3"
/codon_start=1
/product="Homolog of odr-2(two) protein 4, isoform b"
/protein_id="AAM15601.1"
/db_xref="GI:20198879"
/db_xref=" WormBase: T12A2.16b"
/translation="MRKKWLTIVLCTLQATFSLVHDSERKSLMERPKSHCFSCA
SFVLPWSQMLHYPKPKFTDRCWQPDGSGIGLVPCCSACFTLVERIDVVSQHGVI
GVIRGMDRLFLGLDDVRNLIUSAYENQVRCHTRKLRLLPFLSGQTDVDTVFCSN
GDFCNEHMLRELSNSFQHSVLVATWILF"
/join(3241..3365,3586..3740,6392..6558,6795..6926)
/gene="hot-4"
/standard_name="T12A2.16a"
/notes="coded for by the following C. elegans cDNAs:
AF324056"
/codon_start=1
/product="Homolog of odr-2(two) protein 4, isoform a"
/protein_id="AAM15600.1"
/db_xref="GI:20198878"
/db_xref=" WormBase: T12A2.16a"
/translation="MRKKWLTIVLCTLQATFSLVHDSERKSLMERPKSHCFSCA
SFVLPWSQMLHYPKPKFTDRCWQPDGSGIGLVPCCSACFTLVERIDVVSQHGVI
RGMCDRLFLGLDDVRNLIUSAYENQVRCHTRKLRLLPFLSGQTDVDTVFCSN
NEHMLRELSNSFQHSVLVATWILF"
/complement(7287..12051)
/gene="T12A2.15"
/complement(join(7287..7364,7412..7520,7644..7735,
7844..7979,8026..8132,9511..9663,9726..9884,9932..10111,
10156..10458,10598..10687,10735..10911,10953..11156,
11304..11499,11832..11919,11982..12051))
/gene="T12A2.15"
/standard_name="T12A2.15a"
/notes="contains similarity to Pfam domain PF00168 (C2
domain); coded for by the following C. elegans cDNAs:
yk194a9.3, yk194a9.5, yk195a9.5, yk818e11.3, yk818e11.5"
/codon_start=1
/product="Hypothetical protein T12A2.15a"
/protein_id="AAC24452.2"
/db_xref="GI:20198877"
/db_xref=" WormBase: T12A2.15a"
/translation="WSQSLVLPVAGSAILSTFTFFLKGNDYFVWVLIILIASVMKS
YLWKRERLIALATALLREVEVIMAGLDLPAWQPDPTERVENLKVHQLWPYVG
EYTKMDFDLPVQKMGPMGNKFRKTRKMDGLDFCRVGGIKVITINVGDRILVD
MDVAYGADFTVSCCGTGMNNIQSGKRLKLLPFLPMWGVSGVSTFLEMPQM
DFNTLGMVWELPGLDIAIRSVNSIQAALCVLPNEIVPLAPVDVDTQLYFPEPDG
VVRILITAEKNLNDISFIKKKSDPYAETQVGSQFKTRTIDDDNLINWNEYPFAG
VDQAGOKLRLTELPDQDQDEBGRSLSDVLKLVQAGTIDKWYPLEGCKHGLHKA
TWNQSLTELRLEQMEAWGQADPHSALLWYDVSADLPPKSKLEPSPFVEV
SLGKETQTPVKVTVNPLFQSKFLFVRHLEGQELKFEAVDGTTRSLSLNIPLT
LLKPEQNEQNMELTILGVHQSPIVITIRIMURSQQSLSHGRNSRLGRFLRS
KHEWKRERADENRGEIQLDFDDLVNQKLALICRDLMTDFKDDQCNYPVSKL
VALDKNVEVFKKPTPAKNTKTRHFDHVEIDINPSDLLNKHKVINVKDNTYGTFFA
KPVLCGLRIEISLNRQLSQRWIPLSVERK"
/complement(join(7287..7364,7412..7520,7644..7735,
7844..7979,8026..8132,8386..8532))
/gene="T12A2.15"
/standard_name="T12A2.15b"
/notes="contains similarity to Pfam domain PF00168 (C2
domain); coded for by the following C. elegans cDNAs:
yk349e10.5"

```


ORGANISM	Drosophila sp.	Db	1127	TACATAGCATACAGCAGCTATCTACTCGCTGGCAGCATTCCTGTCGATCGATCTCGTTT	1186
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	Qy	258	ValGlyPheGlnProIleAArgThrSerGluHisMetAlaLeuValPheValLeu	277
AUTHORS	Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.	Db	1187	GTGGGATTCCAAACCATCCAGAGCTCCGAACACATGCTGGCAGCTGGGAACCTTTGGCGCTG	1246
TITLE	Detection kits, such as nucleic acid arrays, for detecting the expression of 10,000 or more Drosophila genes and uses thereof	Qy	278	LeuMetAlaValAlaThrIleuAArgHisLeuGlnSerValLeuSerArgAsnGluPheArg	297
JOURNAL	Patent: WO 01/042-A 1433 27-SEP-2001; PE Corporation (NV) (US)	Db	1247	TGCAGATTTCAGCTTTTCGTCGACTATTCGCTCGCATTCGCGCATCCCAAGATCACTTCAT	1306
FEATURES	Location/Qualifiers	Qy	298	LysLeuPhe-----IleValGlyLeuLeuValGlyValGlyValPheValAlaVal	315
source	1..2855 /organism="Drosophila sp." /mol_type="unassigned DNA" /db_xref="taxon:7242"	Db	1307	CTGCTCTTCAAGAGCTGGTTTCCAGTGTTTG-----ACTGTGGTCTCGCTGGGT	1360
ORIGIN		Qy	316	ValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrp	335
Alignment Scores:		Db	1361	ACCTGCTCACGCTTACCGGAAAGTCTCCCTCGCAGCGGAGATTTTACTCGCTACTA	1420
Pred. No.:	2,218-168 Length: 2855	Qy	336	AspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnPro	355
Score:	2226.00 Matches: 417	Db	1421	GATCCATCTATGCCAAGAATCACATTCATCATCGCTCGTATCGAGCATCAGCT	1480
Percent Similarity:	74.20% Conservative: 115	Qy	356	ThrThrTrpPheSerPhePheAspLeuHisIleLeuValCysAlaPheProValGly	375
Best Local Similarity:	58.16% Mismatches: 145	Db	1481	ACATCTGCTGCTCTTTCTATTTCATCTGCAGATCTCTGCTCTCTCTCTCTCTCTCTCT	1540
Query Match:	55.02% Indels: 40	Qy	376	ValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIle	395
DB:	6 Gaps: 8	Db	1541	CTTTATTTCTGCTTTCCAGCTGACGAGTCAACATCTTTATTATTATGTCGCTC	1600
US-10-028-384-8 (1-774) x CQ573675 (1-2855)		Qy	396	SerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrIleuThrProValVal	415
Qy	18	Db	1601	ACCAGTATTATTTCGCGAGTGATGTCGCTTAATGTCGCTCTGCGCGCGCTGATG	1660
Db	467	Qy	416	CysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeuGlnGlu	433
Qy	38	Db	1661	TGTGCTATCCGGAATTCCTATTCGATCTGTCGCGCAAGTATATCAAGAGCTTGAT	1720
Db	527	Qy	434	---AspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGluAla	452
Qy	58	Db	1721	GCCGGAAGCTCATCGAAGCAGGGA-----	1744
Db	587	Qy	453	GluAspSerIleGlnLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArgThr	472
Qy	78	Db	1745	-----GTAGAAAGCAGCGCGCAGCACAAGACTG-----	1774
Db	647	Qy	473	LysHisAspAlaGlnAspThrGlyValSerSerAsnLeuLysSerIleValIleLeu	492
Qy	98	Db	1775	-----GAACAGCAGCAGCGGAGCGGTGAAGAGCGAGTGGCCATTGGATTCTGGGT	1825
Db	707	Qy	493	AlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAla	512
Qy	118	Db	1826	GTCATCACTGATGCTTATAGTGTACACGCTGCACTGACCTGGGTCACTCGGAGGCC	1885
Db	767	Qy	513	TyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIle	532
Qy	827	Db	1886	TATCTCTCGCCAGTATTGTTTGAGT---GCCAGTCCACGATGGCGCGCATCAATT	1942
Db	138	Qy	533	LeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAspAlaArg	552
Qy	158	Db	1943	TTGATGACTTCGCGAGGCTACTACTGCTGTCAGATGAACACTCCGAGCAGCTCGC	2002
Db	886	Qy	553	ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeu	572
Qy	177	Db	2003	ATATGTCCTGGTGGGACTACGCTACCGATTAACCGCCATGCCAATCGGAGATATG	2062
Db	946	Qy	573	ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer	592
Qy	197	Db	2063	GTGGATAACAATACTTGAAACAACACACATATATTCGCGGCTTTCGCGCAGCTGGCTCT	2122
Db	946	Qy	593	ThrGluGlySerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIle	612
Qy	178	Db	2123	TCGAGGAGGAAGCCTACGAGATATGAGGAACTGGATGTGACTACGTTCTCTCGTATT	2182

QY 57 AlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArg 76
DB 234 ACAACAAAGGTTCTTCGCCGAGGAAGGATTTATATAATTCCTAACTGGTTGATGATCGA 293
QY 77 AlaTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
DB 294 GCATGGTATCCATTGGGAGAAATTATAGGAGGAACAATCTATCCAGGCTCATGATACCC 353
QY 97 SerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIle 116
DB 354 TCTGCAGCATCTACCAATTTCTACATCTTTTCCACATCACTATTGACATCCGAATGTT 413
QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
DB 414 TGTGTATCTCTGCTCCATCTCTCTGCTTTCCACACCATAGTCACATTCATCTGACA 473
QY 137 LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaIleSerPheIleAlaIleValPro 156
DB 474 AAAGAGCTGAAAGATGCTGCTGAGGCTTTTGGCAGCAGCAATGATAGCAGTGGTGCCT 533
QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
DB 534 GGCATATATACATCTCTGCTGCTGATCCTATGATATGAGGATTTGCCATTTTCGT 593
QY 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
DB 594 ATGCTTTCTAACTTACTTACATGTTGATTAAGGCAGTAAGACTGTTCAATCTAATGGGCA 653
QY 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyValPheIle 216
DB 654 GCCATGCGCTCTAGCATATTTCTATATGCTCTCTGCTGGGAGGATACGTGTTCTG 713
QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg 236
DB 714 ATCAACCTTATACCATTTCTATGCTGTTTGTATGCTTACCGAAGATTTCTTCATCGA 773
QY 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
DB 774 ATTATATGCTGCTTACTGTCACAGTTTATTTGTTGGCACCATCTCTCCATGCAATCTCC 833
QY 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
DB 834 TTTGTTGGATTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 893
QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
DB 894 CTGTGTCCAGATCCATGCTTTGTAGACTACCTCGGAGCAAGCTCAATCTCCAGCAATTT 953
QY 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
DB 954 GAAATCTTTTAGGAGTGTGATTTCCACTGGTTGGATTTGTTCTTCTACTGTCGGA--- 1010
QY 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
DB 1011 -----GCTGCTCTCATGCTGACC-----GGCAAAATATCCCATGAGTGGTCTTC 1058
QY 332 TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
DB 1059 TATTCCTCTTTGATCCTTTATGCAAAAGACAATATTCCAATCAATGCTTCTGTATCT 1118
QY 352 GluHisGlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla 371
DB 1119 GAGCATCAGCCACGACATGCTCTTCTTATTTTGTATTTGATTTGACGTTACTTCTTCTCATG 1178
QY 372 PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
DB 1179 TTTCCAGTTGGTCTGTATTTCTCTCAGTAATCTATCTGATGCCAGAAATTTTCATATTC 1238
QY 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
DB 1239 ATGTACGGTGTGACCATGTACTTTCTGTCAGTCTATGCTGGTCTGATGCTGCTTCT 1298
QY 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431

DB 1299 GCCCTGTTATGTGTATTTCTTCTGCAATGGGGTTTCTCAGGTATTATCCACATTACATG 1358
QY 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
DB 1359 AAGAATCTTGACATAGCCGCTCT----- 1382
QY 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
DB 1383 -----GATAAAAACA----- 1394
QY 472 ThrLysHisAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
DB 1395 -----AAAAAGCACACAGGATTTCTACATATCCATCAAAACGAAAGTTGCCAGTGGC 1445
QY 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
DB 1446 ATGATCCTGGTCATGGCTTTCTTTTGGTCATTTATACATTCCTCTCTACATGGGTGACA 1505
QY 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
DB 1506 AGTGAAGCTTACT 1562
QY 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAlaAsp 549
DB 1563 AGGATCATCTTGTATGATTTTCAGGGAGGCTTATTTGCTGAGCCGCTGAG 1622
QY 550 AspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
DB 1623 GATGCAAAAGTAAATGTCGTTGGGATTTATGTTACCATGATTCAGCAATGCCAATCGA 1682
QY 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
DB 1683 ACAATCTCTGCTGATAATAATACCTGGAATAACACACACATTTCCAGAGTAGCCGAGCA 1742
QY 590 MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
DB 1743 ATGGCTCCACTGAGAAAGAAAGCCTATGATATTTAGAGAGCTGGATGTTAGTTATGTC 1802
QY 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 629
DB 1803 CTTGTTATTTTGGGGCTCTCAGGATATTCATCAGATGATATCAACAAATTTTGTGG 1862
QY 630 MetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluSer 645
DB 1863 ATGCTAAGAAATGGAGGAAGCACTGACCTGTTAAACAC-----ATCAAGAGCAC 1913
QY 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
DB 1914 GATTACTACACTCTACTGGAGAGTTTCTGTAGATAGAGAAGGATCTCCCGTTCTGCTC 1973
QY 666 AsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArg 685
DB 1974 AACTGCTTATGATACAGATGCTGCTATATCGCTTTGGTCAAGCTTACACCGAACCA 2033
QY 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu 705
DB 2034 CGCCCTCCAGTTATGACAGAGTGAGAACGCGCAAAATCGCAATAAAGATTTCCAGCTT 2093
QY 706 ThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLys 725
DB 2094 GATGTTCTAGAGGAAGCTTACACACAGAGCACTGCTGCTGTGAGATAATATAAAGTAAA 2153

Search completed: December 15, 2004, 06:00:01
Job time : 9328.93 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:28:05 ; Search time 892.033 Seconds
(without alignments)
4554.822 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 4046

Sequence: 1 MNRTPKMLNSKVAGYSSLIT.....RRKGYIRNPVVVKGRITLK 774

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2 1/USPTO.spool/US10028384/runat 14122004 131518 8007/app query.fasta_1.3740
-DB=N Geneseq 23Sep04 -OFMT=fastap -SUFFIX=ring -MINMATCH=0 1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=3 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=ptc -THR_WAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=exat 14122004 131518 8007 -NCPU=6 -ICPU=3
-USER=US10028384@cgn 1.1.1986 -runat 14122004 131518 8007 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

N_Geneseq_23Sep04: *
1: Geneseqn1980s: *
2: Geneseqn1990s: *
3: Geneseqn2000s: *
4: Geneseqn2001as: *
5: Geneseqn2001bs: *
6: Geneseqn2002as: *
7: Geneseqn2002bs: *
8: Geneseqn2003as: *
9: Geneseqn2003bs: *
10: Geneseqn2003cs: *
11: Geneseqn2003ds: *
12: Geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4046	100.0	2417	10	ADD94789
2	4046	100.0	2699	4	ABL13247 Drosophila
3	3918.5	96.8	4922	4	ABL13246 Drosophila
4	2949	72.9	2481	10	ADD94783
5	2948	72.9	2710	10	ADD94785 Mouse SIM
6	2226	55.0	2855	4	ABL02795 Drosophila

7	2198.5	54.3	3094	10	ADD94791
8	2177.5	53.8	2472	10	ADD94793
9	2177.5	53.8	2472	10	ADH28844
10	2176	53.8	2256	6	AB232036
11	2172.5	53.7	2232	8	ABT20789
12	2168.5	53.6	2760	5	ABV24502
13	2113.5	52.2	2466	10	ADB69900
14	2027.5	50.1	2733	10	ADD94787
15	2014	49.8	2603	8	ABT17781
16	2014	49.8	3969	8	ABT17781
17	2014	49.8	4603	8	ABT19595
18	1978	48.9	6153	4	ABL02794
19	1961	48.5	2000	10	ACC61293
20	1888.5	46.7	1848	8	ABT18969
21	1851.5	45.8	2882	10	ADB69539
22	1851.5	45.8	4738	10	ADB69178
23	1840.5	45.5	1969	8	ABT18375
24	1597	39.5	1664	4	AAK94164
25	1597	39.5	1664	12	ADL30661
26	1512	37.4	1543	2	AX85055
27	1512	37.4	1543	8	ACD18981
28	1512	37.4	1543	12	ADG78372
29	1512	37.4	1543	12	ADN60663
30	1509.5	37.3	1209	6	ABL33264
31	1509.5	37.3	1209	6	ABL89850
32	1331	32.9	1371	12	ADP28508
33	1191	29.4	2510	5	AAF93772
34	1191	29.4	2546	2	AAV44866
35	1191	29.4	2546	5	AAF98463
36	1152.5	28.5	2537	4	AAD08289
37	1147.5	28.4	2660	6	ABQ54750
38	1147	28.3	2547	4	AA08315
39	1122	27.7	2284	4	AAH18021
40	1098	27.1	2953	4	ABU18224
41	1006.5	24.9	1114	4	AAH99794
42	963	23.8	787	4	AAH07526
43	923.5	22.8	764	5	AAF93968
44	851.5	21.0	2785	4	ABL18208
45					

ALIGNMENTS

RESULT 1

ADD94789

ID ADD94789 standard; DNA; 2417 BP.

AC ADD94789;

DT 29-JAN-2004 (first entry)

DE Drosophila melanogaster STT3 gene sequence.

KW source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide;
KW fruit fly; gene; ds; STT3.

OS Drosophila melanogaster.

FN WO2003054008-A2.

XX 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

XX 20-DEC-2001; 2001US-00028384.

XX (COMP-) COMPATIGENE INC.

PA

XX Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 DR P-PSDB; ADD94790.
 XX
 XX New human source of immunodominant MHC-associated peptide: (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 XX Claim 6; SEQ ID NO 7; 66pp; English.
 PS
 XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the Drosophila melanogaster SPT3 gene which
 CC is related to the invention. Note: This sequence does not appear in the
 CC specification but was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 2417
 Score: 4046.00 Matches: 774
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0
 US-10-028-384-8 (1-774) x ADD94789 (1-2417)
 QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 DB 71 ATGAATCGGACGCGGAGATGCTGAACAGAGGTGGCTGGGTACAGAGCGCTAATCACC 130
 QY 21 PheAlaTleLeuLeuIleAlaTTPLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 DB 131 TTGGCATCTGCTAATCGCTGGCTGGCGGGGATTTCTCTCGCTCTTCGCGGTCATC 190
 QY 41 ArgPheGluSerIleIleHisGluPheAspProTTPPheAsnTyrArgAlaThrAlaTyr 60
 DB 191 CGTTTCAGTCGATTATCCATGATGTTGATCCGCTGTTCAACTACCGGCCACCGCCTAC 250
 QY 61 MetValGlnAsnGlyTTPTrpAsnPheLeuAsnTTPPheAspGluArgAlaTTPTrpPro 80
 DB 251 ATGGTCAGAAATGGTGGTACAACTTCCCTCACTGGTTTCGACGAGCGCCATGGTATCCG 310
 QY 81 LeuGlyArgTleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 DB 311 CTCGGCAGGATTGTGGCGGTACCGTCTATCCCGGCTGATGATTACGTCCGCGGGAATC 370
 QY 101 HistTrpLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 DB 371 CATGGCTGTGCACCGTACTCAACATACCGGTCCATATTCTGACATCTGCGGTGTTCCCG 430
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTTP 140
 DB 431 GCGCCGATCTTCAGTGGCCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 490
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 DB 491 TC CGGCGCGCGCGCTCTTCGCGCGCAGCTTCATCGCATCTGCTGGCTGTACATCAGT 550
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 DB 551 AGGTCCGCTGGCTGGATCGTACGATACAGGAGCATTCGCTATTCGCCCTCGAGTTACCC 610
 QY 181 TyrPheLeuTTPValArgSerValLysThrGlySerValPheThrPheSerAlaAlaAla 200
 DB 611 TACTTCTCTGGTGGTGGCTCAGTGAAGACTGATCCGCTGTTCTGCTGCGCGCAGCCGCT 670
 QY 201 LeuSerTyrPheTyrMetValSerAlaTTPGlyTyrValPheIleIleAsnLeuIle 220
 DB 671 TTGTCTCTACTTCTACATGTGTCCGCTGGGTGGTGTACGTGTTTCATCAACCTGATA 730
 QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
 DB 731 CCTCTGCAGCTTCTCGTACTGCTCATTTATGGGAGGTACTCGCCGGTCTGTGACACG 790
 QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 DB 791 TACAGCACCTTCTACATCTGGGACTGCTGTTCTCCATGCAGATCCCTTCGTGGGATTC 850
 QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 DB 851 CAACCCGATACGACACAGTGAACATGGCTGGCTGGGAGTGTGTTGCTCTCTGATGGCC 910
 QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 DB 911 GTGCCACCTTGGCCATTTGGCAGTCCGTGCTGTCGCGACAGTTCGGGAGGTGTTTC 970
 QY 301 IleValGlyLeuLeuValGlyValGlyValPheValAlaValValValLeuThrMet 320
 DB 971 ATCGTCGGCGGATGTGCTGGGCGGTGGCTGTTGTCGCGCTGTGCTGTCTACCATG 1030
 QY 321 LeuGlyValValAlaProTTPSerGlyArgPheTyrSerLeuTTPAspThrGlyTyrAla 340
 DB 1031 CTGGGCGTGTGGCCCGGTGGAGTGGAGCGCTTCTACTCGCTGGGATACTGGCTACGCC 1090
 QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 DB 1091 AAGATCCACATTCCTCCATCTGATCCGTGTCGGAGCATCAGCCACCACTTGGTTCG 1150
 QY 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTTPTrpCysIle 380
 DB 1151 TTCTTCTTGTGATCTGCACATCTCTGGTGGCTTCCAGTGGAGTGGTACTGTCATC 1210
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 DB 1211 AAGCAGATCAACAGCAGCGCGTTCCTGCTGGTGTCTGTACGCCATCATCGCTTACTTC 1270
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 DB 1271 GCTGGTGTGATGGTGGCTTTGATGTTGACCTTCACCGCGGTGTGTGTCATGCTGGCGGA 1330
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 DB 1331 GTGGCCCTTTTCGGGACTGTTGGATGTCTCTCAAGAGAGATTCGTCTAAGCGAATGGGC 1390
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
 DB 1391 ACAGCCATAAGCGCAGCCACCGAAGTGGATGAAGTGAAGATTCCTCATTTAGAGAGAGACG 1450
 QY 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThr 480
 DB 1451 CTGTACCAAGGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAGCAGGATACT 1510
 QY 481 GlyValSerSerAsnLeuLysSerIleValIleIleAlaValLeuMetLeuMetMet 500
 DB 1511 GGCTGAGCTCCACCTGGAAGATGTTGTTATTTGGCCGCTTCTAATGCTGTTGATGATG 1570
 QY 501 PheAlaValHisCysThrTTPValThrSerAsnAlaTyrSerSerProSerIleValLeu 520

QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyLeuAlaAlaPheAlaLeuGlnPheThr 180
 DB 560 AGGTCGGTGGCTGGATCGTACGATAACAGAGGCAATGGCCATATTCGCCCTGCAGTTACC 619
 QY 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAla 200
 DB 620 TACTTCCTGTGGGTGGCTCAGTGAAGACTGGATCCGTTCTGTCGGCCGACCGCT 679
 QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle 220
 DB 680 TTGCTCTACTTCTACATGGTGTCCGCTGGGTGGCTACGTGTTTCATCATCAACCTGATA 739
 QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
 DB 740 CCCCTGCACGCTTCCTGACTGCTCATATGGGAGGTACTCGCCGGCTGTGCTGACCCAGC 799
 QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 DB 800 TACAGCACCTTCTACATCTGGGACTGCTGTTCTCCATGCAGATCCCTTCGTGGGATTC 859
 QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 DB 860 CAACCGATACACACAGTGAACATGGCTGCGCTGGAGTGTGTTGCTCTCTATGGCC 919
 QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 DB 920 GTGGCCACCTTGGCCCAATTTGCAGTCCGTCTGTCGCGCAACGAGTTCGGGAAGCTGTT 979
 QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 DB 980 ATCGTCGCGGATTTGCTGGGTGGGTGGCTGTTGTCGCGCTGCTGCTGCTACCATG 1039
 QY 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 DB 1040 CTGGGCGTGTGGCCCGTGGAGTGGACGCTTCTACTCGCTGTGGATACCTGGCTACGCC 1099
 QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrTrpPheSer 360
 DB 1100 AAGATCCACATTCCTCATATGTCATCGCTGCGGAGCATCAGCCACCATCTGGTCTCG 1159
 QY 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
 DB 1160 TTCCTCTTTGATCTGCACATCTGCTGGTGGCTTCCAGTGGAGTGTGGTACTGCATC 1219
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 DB 1220 AAGCAGATCAACGACGAGCGCGTTCCTGCTGCTGTACGCCATCAGTCGGTTCCTTACT 1279
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 DB 1280 GCTGGTGTGATGGTGGCTTTGATGTTGACCTCACGCCGCTGGTGTGATGTCGCGCA 1339
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnAspSerSerLysArgMetGly 440
 DB 1340 GTGGCTTTTCGGGACTGTGGATGTCTCTCAAGAGATTCGTCCTAAGCGAATGGGC 1399
 QY 441 ThrAlaIleSerAlaAlaThrGluValAlaGluAspSerIleGluLysIleThr 460
 DB 1400 ACAGCCATAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTCATTCAGAGAAGAGC 1459
 QY 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThr 480
 DB 1460 CTGACGACAGGCTGCGACGCTGAAGCATGCTAAGCATGATGATCCCGCAGGATGACT 1519
 QY 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet 500
 DB 1520 GGGCTCAGCTCCACCTGAAGAGTATGTTATTTTGGCGCTTCTAATGCTGTGTGATG 1579
 QY 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
 DB 1580 TTCCTGTCCACTCAGCTGGTGACGAGCAATGCTTACTCCAGTCCCTCCATTTGCTGTG 1639
 QY 521 AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspPheArgGluAlaTyr 540

DB 1640 GCTTTCCACAAACAGTCAAGATGGATCCCGCAACATTTAGACGATTTTCAGAGAGCTTAC 1699
 QY 541 TyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGly 560
 DB 1700 TACTGGCTTTCGACAGAACACTCCGATGATCTCGTATGTCTTGGTGGATTCGGA 1759
 QY 561 TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 580
 DB 1760 TACCAGATAGCGGAATGCAACAGAACAGACGCTAGTGGGATAATAAATACGTGGAACAAT 1819
 QY 581 SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle 600
 DB 1820 AGTCATAGCTGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTACGAAT 1879
 QY 601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620
 DB 1880 ATGACATCTCTTGACGTGGACTACGTTTGGTGTGATCTTTGGCGGTGATCGGCTATCT 1939
 QY 621 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 640
 DB 1940 GCGATGATATCAACAAGTTCCTGGGATGGTCCGAATGCTGAGGAGAGCATCCCAAG 1999
 QY 641 AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly 660
 DB 2000 GACATTAAAGAAAGCGATTACTTTACCGACCGCGTGAATTCAGGGTAGATGCCGAAGT 2059
 QY 661 AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680
 DB 2060 GCTCCCGGCTGCTCACTGCTTATGACAAATTAAGTACTACAGATTCGGGAATG 2119
 QY 681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
 DB 2120 AAGTTGGACTACAGAGGTCATCTGATGATGCACACGTAACGCGCTCAITGGGAAT 2179
 QY 701 LysAspPheAspLeuThrTyrLeuGluAlaValTyrThrGluHisTrpLeuValArg 720
 DB 2180 AAGGATTCGATCTGACCTACCTGGAGGAGGCTTACACACAGAACCACTGGCTTGTTCG 2239
 QY 721 IleTyrArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
 DB 2240 ATCTATAGGTTGAAGAAGCGCATGAGTCAATAGACCATCACTGAAGACCAAGAGAGA 2299
 QY 741 ThrIleProAlaAsnPheIleSerArgLysAsnSerLysArgLysGlyTyrIle 760
 DB 2300 ACATTTCTCCAGCAAACTTCAATTCGAGAAGAAGAACTTAAGCGTCCAGGCTACATA 2359
 QY 761 ArgAsnArgProValValValLysGlyLysArgThrLeuLys 774
 DB 2360 CGAAACCGACCGGTGTGTTAAGGGAAGAAACCAACCTTGAAA 2401

RESULT 3

ABL13246/c
 ID ABL13246 standard; cDNA; 4922 BP.

AC ABL13246;
 AC ABL13246;

DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.

OS Drosophila melanogaster.
 OS Drosophila melanogaster.
 OS Drosophila melanogaster.
 OS Drosophila melanogaster.

XX WO200171042-A2.
 XX WO200171042-A2.

XX 27-SEP-2001.
 XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.
 PR 23-MAR-2000; 2000US-0191637P.

```
PR 11-JUL-2000; 2000US-00614150.
XX (PEXE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EM;
XX WPI: 2001-656860/75.
XX P-PDSB; ABB69143.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX
XX Claim 1; SEQ ID NO 34220; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signaling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX CC sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX
XX Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 0 Length: 4922
XX Score: 3918.50 Matches: 772
XX Percent Similarity: 90.93% Mismatches: 0
XX Best Local Similarity: 90.93% Indels: 77
XX Query Match: 96.85% Gaps: 4
XX DB: 4
XX
XX US-10-028-384-8 (1-774) x ABL13246 (1-4922)
XX
XX QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
XX DB 3843 ATGAATCGAGCCGCGAAATGCTGAACAGCAGAGGTGGCTGCTACAGCAGCCTAATCACC 3784
XX
XX QY 21 PheAlaIleLeuLeuIleAlaThrLeuAlaGlyPheSerArgLeuPheAlaValIle 40
XX DB 3783 TTCGCCATCTCTAATCGCTGGCTGGCGGGATTTCTCTCGCTCTTCGCGGTATC 3724
XX
XX QY 41 ArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArgAlaThrAlaTyr 60
XX DB 3723 CGTTTCGAGTCGATATCCATGAGTTTGTATCCGTGGTTCAACTACCGGGGCCACCGCTAC 3664
XX
XX QY 61 MetValGlnAsnGlyTyrPtyrAsnPheLeuAsnTrpPheAspGluArgAlaTyrPro 80
XX DB 3663 ATGGTGCAATGGTTGGTACAACTTCCTCAACTGGTTTCGACGAGCGCGATGTATCCG 3604
XX
XX QY 81 LeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
XX DB 3603 CTCGGCAGGATTTGTGGGGGTACCGTCTATCCCGGCTGTATGATTCGCGGGAATC 3544
XX
XX QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
XX DB 3543 CATTCGCTCTGCAGTACTCAACATACCGTCCATATTCGTGACATCTGCGTGTCTCTG 3484
XX
XX QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
XX DB 3483 GGGCGCATCTTCAGTGGCCGTACCTCCATCTCCACTACCTGCTGACCAAGGAGCTGTG 3424
XX
XX QY 141 SerAlaGlyValAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
XX DB 3423 TCCGGGGCGCGGCTCTCTCCGCGCCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT 3364
XX
XX QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
XX DB 3363 AGTCGGTGGCTGGATCGTACGATAACGAGGGCATTTGCCATATTTCGCGCTCGAGTTCA 3304
XX
XX TyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSerAlaAlaAla 200
XX DB 3303 TACTTCTCTGTGGTGGCTCAGTGAAGACTGGATCCGTGTTCGTGGTGGCGGCGAGCGCT 3244
XX
XX QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIle 220
XX DB 3243 TTGTCTTACTTCTACATGGTGTGCGCTGGGGTGGCTACGTGTTCATCATCACCTGATA 3184
XX
XX QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
XX DB 3183 CCCCTGACGCTTCCTGACTGCTCATTTATGGGAGGTACTCGCGGCTCTGCTGACACG 3124
XX
XX QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
XX DB 3123 TACAGCACCCTTCTACATCTCTGGGACTGCTGTTCCTCATGCAGATCCCTTCGTGGGATTC 3064
XX
XX QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
XX DB 3063 CAACCGATATCAGCAGCAGTGAACACATGGCTGGGAGTGTTCGTCTCTTATGGCC 3004
XX
XX QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
XX DB 3003 GTGGCCACCTTGGCCCATTTGCAGTCCGTCTGTGGCGCAACGAGTTCGGAAGCTGTTTC 2944
XX
XX QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuLeuMet 320
XX DB 2943 ATCGTCGGCGGATTCGCTGGTGGCGGTTCGCTCTTTGTGGCCGCTCGTGGTCTCACC 2884
XX
XX QY 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
XX DB 2883 CTGGCGGTGTGGCCGCTGGAGTGGAGCGCTTCTACTCGCTGTGGGATCTAGCTACGCC 2824
XX
XX QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrTrpPheSer 360
XX DB 2823 AAGATCCACATTCGCATCATTCGCTGTGCGAGCATCAGCCACCATCTGCTGTCTCG 2764
XX
XX QY 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrCysIle 380
XX DB 2763 TTCTTCTTTGATCTGCACATCTCTGTGTGGCCCTTCCACGTGGAGTGTGTACTGCATC 2704
XX
XX QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
XX DB 2703 AAGCAGATCAACGACGAGCGGCTTTTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2644
XX
XX QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
XX DB 2643 GCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2584
XX
XX QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
XX DB 2583 GTGGCCCTTTTCGGGACTGTTCGATGTTCCTGTGCAGAGGATTCGTCTAAGCGAATGGGC 2524
XX
XX QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
XX DB 2523 ACAGCCATTAACGCGCAGCCACCCAGAGTGGATGAGCTGAGGATTCATTTGAGAAAGAGCG 2464
XX
XX QY 461 LeuTyrAspLysAla----- 465
XX DB 2463 CTGTACGCAAGGT- GAGTTCTTACTAACACATCCATCCGATTTGATTTTAAATATACA 2405
XX
XX QY 466 -----GlyLysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThr 480
XX DB 2404 TCGCATTTGCGAGCTGGCAAGCTGAGCATCTGACTAAGCATGATGCGCCAGCAGGATACT 2345
XX
XX QY 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet 500
XX DB 2344 GCGCTCAGCTCCAACTCAAGAGTATTGTTATTTCGCGCTTCTTAATGCTGTTCATGATG 2285
XX
XX QY 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
XX DB 2284 TTCGCTGTTCCTACTGACGCTGGGTGACGACGACATCCCTACTCCAGTCCCTCCATTTGCTTG 2225
```

QY 521 AlaPheHisAsnSerGlnAspGly 528
 DB 2224 GCTTTCCACACAGTCAGATGGTAAAGCACCAGTAGCATCATGCTCTCTTTAAACCTA 2165
 QY 529 -----SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyr 541
 DB 2164 TCTAACCTTTTCTCCACTAGATCCGCAACATTTTAGACGATTTTCAGAGAGCTTACTATC 2105
 QY 542 TrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGlyTyr 561
 DB 2104 TGGCTTTCCGAGAACACTGCGGATGATGCTCGGTTATGCTTGTGGGATACCGATAC 2045
 QY 562 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 581
 DB 2044 CAGATAGCGGATGGCAACAGACGACCTAGTGGATAAATAACGTGGAAACATAGT 1985
 QY 582 HisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIleMet 601
 DB 1984 CACATAGCGCTGTTGGCAGGCAATGCTTTCACCGAGGAGAGCTCTACCAATATATG 1925
 QY 602 ThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 621
 DB 1924 ACATCTCTGACGTGGACTACGTTTGGTGTATCTTTGGCGGTGTGATCGGCTATTCTGGC 1865
 QY 622 AspAspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAsp 641
 DB 1864 GATGATATCACAGTTCTTGTGATGGTCCGAAATGCTGAGGAGAGAGATCCAGAGAC 1805
 QY 642 IleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAla 661
 DB 1804 ATTAAAGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGCTAGATCCGAGAGTCT 1745
 QY 662 ProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLys 681
 DB 1744 CCGGCTCTGCTCACTGCTTATGTACAAATTAAGTACTACAGATCGGGAAATGAG 1685
 QY 682 LeuAspTyrArg 685
 DB 1684 TTGACTACAG-GTAAAGCGAAACATTTCTCAGGTAGCGATGACACTAACTCCCTGT 1626
 QY 686 -----GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
 DB 1625 TTCAGAGTTCATCTGATATGATCGCACACGTACCGCGTATGCGGATTAAGACTTC 1566
 QY 704 AspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisThrLeuValArgIleTyrArg 723
 DB 1565 GATCTGACCTACCTGGAGGAGGCTACCCACAGAACACTGGCTGTTCGCATCTATAGG 1506
 QY 724 ValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIlePro 743
 DB 1505 GTGAGAGCGCGATGATTCATAGACCATCATCTGAAGACCAAGAGAGACGATTCCT 1446
 QY 744 ProAlaAsnPhelleSerArgLys 751
 DB 1445 CCAGAAACTTCATTTCGAGAAAGTAGTGATCCCTGACGCTCCCAATGAATTCATTAA 1386
 QY 752 -----AsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765
 DB 1385 CAATCTGATCTTTGACAGACTCTAGCGTCGCAAGGGTATACATAGCAACCGACCGTT 1326
 QY 766 ValValLysGlyLysArgThrLeuLys 774
 DB 1325 GTTCTTAAGGAAAAACGAACCTTGAAA 1299
 RESULT 4
 ADD94783
 ID ADD94783 standard; cDNA; 2481 BP.
 XX
 AC ADD94783;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human SIMP cDNA sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
 KW gene; ss.
 XX Homo sapiens.
 OS
 XX WO2003054008-A2.
 FN
 XX 03-JUL-2003.
 PD
 XX 18-DEC-2002; 2002WO-CA001967.
 XX
 XX 20-DEC-2001; 2001US-00028384.
 PR
 XX (COMP-) COMPATIGENE INC.
 PA
 XX Perreault C, McBride K;
 PI
 XX WPI; 2003-559122/52.
 DR
 XX P-PSDB; ADD94784.
 DR
 XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 PS Claim 6; SEQ ID NO 1; 66pp; English.
 XX
 CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcoma, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,46e-291 Length: 2481
 Score: 2949.00 Matches: 556
 Percent Similarity: 83.88% Conservative: 89
 Best Local Similarity: 72.30% Mismatches: 112
 Query Match: 72.89% Indels: 12
 DB: 10 Gaps: 7
 US-10-028-384-8 (1-774) x ADD94783 (1-2481)

QY 10 SerLysValAlaGlyTyrSerSerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeu 29
 DB 187 TGCAGCCGCTGGGTGGAGTGGCTTCTCTCTTCCATCTCTCTCTGGGCTGCTT 246
 QY 30 AlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 49
 DB 247 GCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTCGAAGCATCATCCAGAGTTC 306

QY	50	AspProTrpPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPhe	69
DB	307	GACCCGTTGGTTAACTATAGATACACATCATCTGTGCATCTCATGGGTTCTATCAATTT	366
QY	70	LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyAcrIleValGlyThrVal	89
DB	367	TTAAATTGGTTTGGTGAAGAGACATGGTATCCCATAGGAAGAATAGTAGGTGGTACTGTT	426
QY	90	TyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIle	109
DB	427	TACCCAGGGTTGATGATACCGCTGGCCTTATTCATTGCATTTTAAATACATTGAACATA	486
QY	110	ProValHisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer	129
DB	487	ACTGTTTCATAGAAGACGATGTGTGTCTCTTGACCACCACTTTTAGGGGGCTTACACPT	546
QY	130	IleSerThrTyrLeuLeuThrIysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAla	149
DB	547	ATATCTACTTCTCTGCTTACAAAGAGAACTTTGGAAACCAAGGAGCAGACTTTTAGCTGCT	606
QY	150	SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn	169
DB	607	TGTTTATTCGTATTTACAGGCTACATATCTCGTCAGTAGTGGATCCCTTTTGATAAT	666
QY	170	GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValIys	189
DB	667	GAAGCATGTCTATTTTGCACCTTCAGTTCACATACTATTATGGTAAATCTGTAAA	726
QY	190	ThrGlySerValPheTrpSerAlaAlaIleLeuSerTyrPheTyrMetValSerAla	209
DB	727	ACTGGTCAGTTTTTTGGACAAATGTCGTGCTTATCTCTATTTCTATATGGTCTCTGCT	786
QY	210	TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIle	229
DB	787	TGGGGTGGTTATGTATTATCATCAATCTTATTCACCTGCATGTATTGTGTGTTTACTG	846
QY	230	MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeu	249
DB	847	ATGCAGAGATACGCAAAAGAGTCTACATGATATAGCATCTTCTCTACATGTGGGTTA	906
QY	250	LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet	269
DB	907	ATATTATCAATGCAGATACCTTTTGTGGATTCGAGCAATCAAGAACAGTGAACACATG	966
QY	270	AlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSer	289
DB	967	GCAGCTGCAGGTGTCTTTTGGATTCCTGCAAGCTTATGCTTCTTTCAGTATCTGAGAGAC	1026
QY	290	ValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyGlyLeuLeuValGlyVal	309
DB	1027	CGATTACAAAAACAAGAGTTCGAGCCCTTTCTTTTGGGTGTATCAGTGCAGGTGCAGGT	1086
QY	310	GlyValPheValAlaValValValLeuThrMetLeuGlyValValAlaProTrpSerGly	329
DB	1087	GCTGTGTCTCTTAGTGTCACTATTTCATATACAGGTTTACATTCGACCACCATGGAGTGGC	1146
QY	330	ArgPheTyrSerLeuTrpAspThrGlyTyrAlaIleIleHisIleProIleIleAlaSer	349
DB	1147	AGGTTTATTTCATGTGGGATCTGGGTATGCAAAAATACACATTCCAATATTATGCATCA	1206
QY	350	ValSerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuVal	369
DB	1207	GTGTCTGAGCATCAACCTACGACTTGGGTGTCITCTTCTTCTTCATCTACATATCTTGTA	1266
QY	370	CysAlaPheProValGlyValTrpTyrCysIleIysGlnIleAsnAspGluArgValPhe	389
DB	1267	TGACCTTCCAGCAGCGCCTTTGGTCTCGATCAAAAATACAAACCATGAACAGATATT	1326
QY	390	ValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu	409
DB	1327	GTTGCTCTATATGCAATCACTGCTGCTACTTTCCTGGAGTGAATGGTGCATGATGTTG	1386
QY	410	ThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspVal	429

Db	1387	ACTTTGACTCCAGTCGTTGATGCTGTCGCAATGCCCTTTCAAATGTTTGGACAC	1446
Qy	430	PhelGluGlnAAspSerSerLysArgMetGlyThrAlaIleSerAlaIaThrGluVal	449
Db	1447	TATTTGGGGATGAC---ATGAAAGGGAAAAATCCACTCTGGAGGACAGCATGATGAG	1503
Qy	450	AspGluAlaGluAAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeu---	468
Db	1504	GAT-----GACAAAGAAACCAAGGAATTTGTATGATAGGAGGAGTAAGTGAGG	1554
Qy	469	LysHisArgThrLysHisAspAlaGlnAAspThrGlyValSerSerAsnLeuLysSer	488
Db	1555	AAACATGCACATGAACGAGAAAAAATCGAAGAG---GGATTAGGCCCTAATAATAAAAGC	1611
Qy	489	IleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpVal	508
Db	1612	ATTGTCACCATGTTGATGCTGATGCTATTGATGATGTTGCTGCTGCAGTCTACTGGGTC	1671
Qy	509	ThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly	528
Db	1672	ACAAGCAATGCCTACTCTAGTCCAAGTGTAGTCTGGCCTCATACAAT---CATGATGCC	1728
Qy	529	SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAla	548
Db	1729	ACCAGGAATATCTAGATGATTTTAGAGAAGCTTACTTTTGGCTAAGCGAAAATACAGAT	1788
Qy	549	AspAspAlaArgValMetSerTyrTrpAspTyrGlyTyrGlnIleAlaGlyMetalAsn	568
Db	1789	GAACATCGACGAGTAATGTTCTGGTGGATTATGGCTATCAGATAGCTGGAAATGGCTAAT	1848
Qy	569	ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys	588
Db	1849	AGAACTACGTTGGTGATATAACACCTGGAATPACAGCCACATAGCATCTGGTGGGAAA	1908
Qy	589	AlaMetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyr	608
Db	1909	GCTATGCTTCTTAATGAAACAGCAGCCCTATAAAATCATGAGGACTCTAGATGTAGATTAT	1968
Qy	609	ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu	628
Db	1969	GTTTTGGTTATTTTGGAGGGTTATTGGCTATTCTGGTGATGATATCAACAAATTTCTC	2028
Qy	629	TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe	648
Db	2029	TGGATGGTTAGGATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAGTGACTATTTT	2088
Qy	649	ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu	668
Db	2089	ACCCACGAGGAGAAATCCGTTGACAAAGCAGATCCCTTACTTTGTTGAATTCGCTT	2148
Qy	669	MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer	688
Db	2149	ATGTATAAAATGTCACTACTACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCCA	2208
Qy	689	GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu	708
Db	2209	GGTTTTCGCGACACGTAATGCTGAGATTTGAAATAAGGACATTTAAATTCAAACATTGT	2268
Qy	709	GluGluAlaTyrThrThrGluHisThrLeuValArgIleTyrArgValLysProHis	728
Db	2269	GAAGAAGCCTTTACATCAGAACACTGGCTTTGTAGATATATAAGTAAAGACCTGAT	2328
Qy	729	GluPheAsnArgProSerLeuLysThrLysGluArgThr-----IleProProAla	745
Db	2329	-----AACAGGGAGACATTAGATCACAAACCTCGAGTCACCAACATTTTCCAAAACAG	2382
Qy	746	AsnPheIleSerArgLysAsnSerLysArgLysGlyTyrIleAcArgAsnArgProVal	765
Db	2383	AGATATTCTCAAGAGAGACTACCAAGAGAGCGTGGCTACATTAAATAAAGCTGTT	2442
Qy	766	ValValLysGlyLysArgThrLeuLys 774	

[illegible]

[illegible]

ID ADD94791 standard; DNA; 3094 BP.
 AC ADD94791;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Mouse ITM1 gene sequence.
 XX
 KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytotoxic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 KW murine; gene; ds; ITM1.
 XX
 OS Mus musculus.
 XX
 PN WO2003054008-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 18-DEC-2002; 2002WO-CA001967.
 XX
 PR 20-DEC-2001; 2001US-00028384.
 XX
 PA (COMP-) COMPATIGENE INC.
 XX
 PI Perreault C, McBride K;
 XX
 DR WPI; 2003-559122/52.
 DR P-PSDB; ADD94792.
 XX
 XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 PS Disclosure; SEQ ID NO 9; 66pp; English.
 XX
 CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytotoxic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the mouse ITM1 gene which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX
 SQ Sequence 3094 BP; 758 A; 651 C; 703 G; 982 T; 0 U; 0 Other;
 XX

Alignment Scores:
 Pred. No.: 2,658-214 Length: 3094
 Score: 2198.50 Matches: 416
 Percent Similarity: 73.78% Conservative: 113
 Best Local Similarity: 58.02% Mismatches: 143
 Query Match: 54.34% Indels: 45
 DB: 10 Gaps: 8

US-10-028-384-8 (1-774) x ADD94791 (1-3094)
 QY 17 SerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeu 36

Db 157 ACACITCTTAAAGCTTCTCATCTGTCGATGGTGTATCTTTTCTACTCGCTT 216
 QY 37 PheAlaValIleAspGlyPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArg 56
 Db 217 TTGCTGCTGAGATTGAAGTGTTCATCCATGAGTTTGCATCCGATTTAAATATCGG 276
 QY 57 AlaThrAlaTrpMetValGlnAsnGlyTrpTrpAsnPheLeuAsnTrpPheAspGluArg 76
 Db 277 ACTACCCGGTTTCTGGCTGAGAGGGGTTTATAAATTCATTAATGTTGATGACCG 336
 QY 77 AlaTrpTrpProLeuGlyValGlyIleValGlyGlyThrValTrpGlyLeuLeuIleThr 96
 Db 337 GCTTGGTACCTTTGGGCGGAATCATTTGGAGGAACAATTTACCAGGTTTAAATGATCAT 396
 QY 97 SerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIle 116
 Db 397 TCTGCTGCAATCTACCATGTACTCTTCTTCATATCACTATTGACATTCGGAATGTC 456
 QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTrpLeuLeuThr 136
 Db 457 TGTGTTTCTTGGCCCACTTTTCTCTCTTCCACCACCATCTTACGTACCACTTACC 516
 QY 137 LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
 Db 517 AAAGAGCTCAAGGATGAGGAGCTGGGCTTCTTGTCTGCCATGATTCGTAGTTCCT 576
 QY 157 GlyTrpIleSerArgSerValAlaGlySerTrpAspAsnGluGlyIleAlaIlePheAla 176
 Db 577 GGGTATATTTCTGATCTGATGCTGGCTCTTATGATAAAGAAGAAATTCATCTTTGTC 636
 QY 177 LeuGlnPheThrTrpPheLeuTrpValArgSerValIleThrGlySerValPheTrpSer 196
 Db 637 ATGTGCTTACTTACTACTGATCAAGGCACTGAGACTGGTTCCTATCTATTGGGCT 696
 QY 197 AlaAlaAlaLeuSerTrpPheTrpMetValSerAlaTrpGlyGlyTrpValPheIle 216
 Db 697 GCCAAGTGTGCGCTCGCTTATTTCTCATGGTCTTTCATGGGAGGCTATGTGTTCTTG 756
 QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlyArgTrpSerProArg 236
 Db 757 ATCAACTTGATTCCTCTACATGCTCTGCTGCTAATGCTGACAGCGCGTTTTCTCACCG 816
 QY 237 LeuLeuThrSerTrpSerTrpPheTrpIleLeuGlyLeuLeuPheSerValGlnIlePro 256
 Db 817 ATCTAGTAGCTACTGATCTTCTGCTGGGCACTTCTTCTATGAGATTCCTCC 876
 QY 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
 Db 877 TTTGTTGTTTCCAGCCGCTCTTTCATCAGAACACATGCGACGCTTTGGAGTGTTCGT 936
 QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
 Db 937 CTCTGTCAGATCCATGCTTTCGTAGATTAATCTCGCAGCAAGTGAATCCACAGCAATTC 996
 QY 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyVal 311
 Db 997 GAAGTTCTTTCCGGAGTGTATCTCCCTGGTGGCTTTGCTCTCTCTCTCTCTCTCTCTCT 1053
 QY 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
 Db 1054 -----GCTCTCTCTCATGCTAAACA-----GGAAAAATTTCTCCCTGGACAGCGCGTTC 1101
 QY 332 TyrSerLeuTrpAspThrGlyTrpAlaLysIleHisIleProIleAlaSerValSer 351
 Db 1102 TACTCTCTGCTGGATCCCTCTTATGCTAAGAATAACATTCCTCCATTTATGCTCTGTTCT 1161
 QY 352 GluHisGlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla 371
 Db 1162 GAGCACCAGGCCACACACTGCTTCTCTACTATTTGATCTACAGCTCTCTCTCTCTCTCTCT 1221
 QY 372 PheProValGlyValTrpTrpCysIleLysGlnIleAsnAspGluArgValPheValVal 391

Db 1222 TTCCAGTGGCTCTATTACTGCTTTAGCAACCTGTCTGATGCTCGGATTTTATCATC 1281
 Qy 392 LeuTyAlaIleSerAlaValTyPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
 Db 1282 ATGTATGGTGTGACCAAGCATGTAATTTTCAAGTGTATGTCGTCTATGCTGTATG 1341
 Qy 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuPheValPheLeu 431
 Db 1342 GCACCTGTATGTCATCTTTTCCGATGTTGTTTCCAGGTCGTCTGCACATATATG 1401
 Qy 432 GluGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
 Db 1402 AAAAATCGACATAGTCCCA----- 1425
 Qy 452 AlaGluAspSerLeuGlyLysThrLeuTyAspLysAlaGlyLysLeuLysHisArg 471
 Db 1426 -----GACAGAGAGCAAG----- 1440
 Qy 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
 Db 1441 -----AAGCAACAGGATTCTACTTACCTTATTAGCAATGAGTGGCGAGTGG 1488
 Qy 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
 Db 1489 ATGATCTGTGTCATGCTTTTCTCATCAGCTGCTTTTCTCATGCTGCTGCTGCTGCTG 1548
 Qy 510 SerAsnAlaTySerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
 Db 1549 AGTCAAGCTTATCTTCTCTCATGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1605
 Qy 530 ArgAsnIleLeuAspAspPheArgGluAlaTyTrpLeuSerGlnAsnThrAlaAsp 549
 Db 1606 AGGATCATTTTGTATGCTTCCAGAGAGGCTATTATGCTGCTGCTGCTGCTGCTGCTG 1665
 Qy 550 AspAlaArgValMetSerTrpTrpAspTyGlyValGlnIleAlaGlyMetAlaAsnArg 569
 Db 1666 GATGCAAAAGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1725
 Qy 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
 Db 1726 ACAATTTTGTGACATATACATGCAATATATACCAATATTTCTCGATGAGGCGAGCA 1785
 Qy 590 MetSerSerThrGluGlyLysSerTyGluIleMetThrSerLeuAspValAspTyVal 609
 Db 1786 ATGCAATCCACAGAGAAAGGCTATGAATCATGAGGAGCTTGATGCTGCTGCTGCTG 1845
 Qy 610 LeuValIlePheGlyGlyValIleGlyTySerGlyAspAspIleAsnLysPheLeuTrp 629
 Db 1846 CTGTGCTATTTTGGAGCCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1905
 Qy 630 MetValArgIle---AlaGluGlyGlyHisProLysAspIleLysGlySerAspTyPhe 648
 Db 1906 ATGCTCGGATGAGGAGAACACAGACAGAGAGAGACATTAAGGAATGACTACTAT 1965
 Qy 649 ThrAspArgGlyLysPheArgValAspAlaGlyAlaProAlaLeuLeuAsnCysLeu 668
 Db 1966 ACTCTACTGGGAATTCGTGTTGATGCTGAGGGTTCTCCGGTCTGCTCAACTGCTT 2025
 Qy 669 MetTyLysLeuSerTyTrpArgPheGlyGluLeuLysLeuAspTyArgGlyProSer 688
 Db 2026 ATGTACAAATGCTTACTACCTTGGGAGGCTCTACACAGAACCAAGGCTCCACCA 2085
 Qy 689 GlyTyAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyLeu 708
 Db 2086 GGCTTTGACCGGTGCGAAATGCTGAGTTGATGATTAATAAGACTTTGACGCTGATGCTG 2145
 Qy 709 GluGluAlaTyThrThrGluHisTrpLeuValArgIleTyArgValLys 725
 Db 2146 GAGGAACGCTATACCAACAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2196

XX ADD94793;
 AC 29-JAN-2004 (first entry)
 DT Human ITM1 gene sequence.
 DE
 XX source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytotoxic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
 KW gene; ds; ITM1.
 XX Homo sapiens.
 OS
 XX WO2003054008-A2.
 PN
 XX 03-JUL-2003.
 PD
 XX 18-DEC-2002; 2002WO-CA001967.
 PF
 XX 20-DEC-2001; 2001US-00028384.
 PR
 XX (COMP-) COMPATIGENE INC.
 PA
 XX Perreault C, McBride K;
 PI
 XX WPI; 2003-559122/52.
 DR
 XX P-PSDB; ADD94794.
 XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX Disclosure; SEQ ID NO 11; 66pp; English.
 PS
 XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytotoxic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human ITM1 gene which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX Sequence 2472 BP; 566 A; 568 C; 583 G; 755 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,69e-212 Length: 2472
 Score: 2177.50 Matches: 415
 Percent Similarity: 72.92% Conservative: 110
 Best Local Similarity: 57.64% Mismatches: 144
 Query Match: 53.82% Indels: 51
 DB: 10 Gaps: 9
 US-10-028-384-8 (1-774) x ADD94793 (1-2472)
 Qy 17 SerLeuIleThrPheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerArgLeu 36
 ID ADD94793 standard; DNA; 2472 BP.

RESULT 8
 ADD94793
 ID ADD94793 standard; DNA; 2472 BP.

Db 152 ACACCTTTTGAAGCTTCTCATCTCTCAATGCTGCTCTATTATTCCTTCTCCACTCGCTG 211
Qy 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 56
Db 212 TTTGCTGCTCCTGAGATTGTGAAGTGTATCCATGAGTTTTCATCGTACITTTATATTCGG 271
Qy 57 AlaThrAlaTyrMetValGlnAsnGlyTyrTrpAsnPheLeuAsnTrpPheAspGluArg 76
Db 272 ACTACCAAGTTCTCGCTGAGAGGAGGGTTTATATAATTCATACACTGGTTGATGACCGA 331
Qy 77 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
Db 332 GCCTGGTACCTTTGGGACGAATCATTTGGAGGAACAATTTACCCAGGTTTATGATCAC 391
Qy 97 SerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIle 116
Db 392 TCTGCTGCAATCTACCAATGTTCTCCATTTTTCACATCACCATCGACATTCGGAATGTC 451
Qy 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
Db 452 TGTGTGTTCTGGCCCT 511
Qy 137 LysGluLeuTrpSerAlaGlyValAlaGlyLeuPheAlaIleSerPheIleAlaIleValPro 156
Db 512 AAAGAGCTCAAGAGTCAAGGAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 571
Qy 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
Db 572 GGATATATCTCCGATCTGTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 631
Qy 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSer 196
Db 632 ATGTCTACTCAGCTACTATCATGTGATCAAGCAGTAAAGACTGGTTCCTCTCTCTCTCT 691
Qy 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 216
Db 692 GCTAAGTGTGCTTCT 751
Qy 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg 236
Db 752 ATCAACTTAATCT 811
Qy 237 LeuLeuThrSerTyrSerTyrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
Db 812 ATCTATGTGGCTACTGCTACTGTTTACTGCTGGGTACTATCTTCTAGGCAGATCTCC 871
Qy 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaLeuGlyValPheVal 276
Db 872 TTTGTGGTTTCCAGCT 931
Qy 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
Db 932 CTCTCCAGATCCATGCTTGTGGATTACTCTGGCAGCAAGTTGAATCCACAACTTT 991
Qy 297 ArgLysLeuPhe-----IleValGlyGlyLeuValGlyValGlyVal 311
Db 992 GAAGTCTTTTCCGAGCGCTCATCTCTCTGTTAGGCTTGTCTCTCTCCCGGGGA--- 1048
Qy 312 PheValAlaValValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
Db 1049 -----GCTCTCTCTCATCTGACA-----GGAAATAATCTCTCTGAGCGGCGCTTC 1096
Qy 332 TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
Db 1097 TACTCAGTGGGATCCCTCTCTATGCTTAAGAACAAACATCCCATCATCTCTCTCTCTCT 1156
Qy 352 GluHisGlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla 371
Db 1157 GAGCATACGCCACAACTGGTCTCTATCTATTTTGCCTGCGAGCTCTCTCTCTCTCTCT 1216
Qy 372 PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
Db 1217 TTTCCAGTTGGCTCTATTACTGCTTTAGCAACCTGCTGATGCCCGGATTTTATCATC 1276

RESULT 9

ADH28844

ID ADH28844 standard; DNA; 2472 BP.

XX

Qy 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
Db 1277 ATGTATGCTGTGACAGCATGTACTTTTCTGAGCTGTAAATGGTGGTCTAATGCTAGTGTG 1336
Qy 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
Db 1337 GCACCTCTGTATGAGCATCTCTCTGCAATGGAGTCTCCAGAGTCTCCACATCATCATG 1396
Qy 432 GlnGluAspSerSerTyrArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
Db 1397 AAGAATCTGGACATAAGTCGCCCA----- 1420
Qy 452 AlaGluAspSerIleGluLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
Db 1421 -----GACAAAGAGGCAAG----- 1435
Qy 472 ThrLysHisAspAlaGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
Db 1436 -----AAGCAACAGGATTCACCTACCTATTAAAGATTGAAGTGGCAAGTGGG 1483
Qy 490 ValIleLeuAlaValLeuMetLeuLeuMetPheAlaValHisCysThrTrpValTrp 509
Db 1484 ATGATCTGTGTCATCGGTTCTTCTCTCATCCTACCTTCTCATCACTTCTCATCACTTCT 1543
Qy 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
Db 1544 AGTGAGGCTTACTCTCTCGTCCATGTACTATCTGCTGCTGCTGCTGCTGCTGCTGCT 1600
Qy 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAsp 549
Db 1601 AGGATCATTTTGTGATCTTCCGAGAGCAATATTTGCTTCTGCTCATATCTTCTCAGAG 1660
Qy 550 AspAlaAlaValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
Db 1661 GATCGAAGTCTATGCTGCTGGATTTATGGTATCAGATTACAGCTATGCGCAACCGCA 1720
Qy 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
Db 1721 ACAATTTTGTGAGCAATAACATGATGATATATCCATATTTCTCGAGTAGGAGGCA 1780
Qy 590 MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
Db 1781 ATGGCGTCCACAGAGGAAAAAGCTATGAGATCATGAGGAGCTCGATGCTAGCTATGTG 1840
Qy 610 LeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 629
Db 1841 CTGCTCATTTTGTGAGGCTTCTCTGCTGATGATATCAACAGTTTCTTCTGCTGCTGCT 1900
Qy 630 MetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluSer 645
Db 1901 ATGTCTCGGATTGGAGGAGCACAGATACAGGCAACAT-----ATCAAGGAGAT 1951
Qy 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGlyAlaProAlaLeuLeu 665
Db 1952 GACTATTATATCTCAACTGGGAGTTCCTGCTGGAGCGTGAAGGTTCTCCAGTGTGCTC 2011
Qy 666 AsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArg 685
Db 2012 AACTGCCTCATGTACAAAGTGTGTACTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2071
Qy 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu 705
Db 2072 CGTCTCCAGCTTTGACCGTTCGGAATGCTGAGATTGGGAATAAGACTTTGAGCTT 2131
Qy 706 ThrTyrLeuGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLys 725
Db 2132 GATGCTCTGGAGGAGGCTATACACAGAACATTTGCTGCTGCTGCTGCTGCTGCTGCT 2191

QY 472 ThrLysHisAspAlaGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
 DB 1436 -----AACCAACAGGATCCACCTACCTATTAAGATTGAAGTGGCAAGTGGG 1483
 QY 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
 DB 1484 ATGATACTGGTCAATGGCTTTCTTCTCATCACTACACCTTTTCAATCAACCTGGTGACC 1543
 QY 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
 DB 1544 AGTGAAGCTACTCTCTCCGTCATTTGCTATCTGCTCTCCCTGGTGGG---GATGGCAGT 1600
 QY 530 ArgAsnIleLeuAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAlaAsp 549
 DB 1601 AGGATCATATTTGATCACTCCGAGAGCATATTTGGCTTCGCTATACTACTCCAGAG 1660
 QY 550 AspAlaArgValMetSerTrpAspTyrGlnIleAlaGlyMetAlaAsnArg 569
 DB 1661 GATGGAGGTCATGCTCTGGTGGATTTGCTATCAGATTACAGTATGGCAACCGA 1720
 QY 570 ThrThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
 DB 1721 ACAATTTTGTGGACAATAACACATGGAAATAATACCATATTTCTCGAGTAGGGCAGCA 1780
 QY 590 MetSerSerThrGluGluLysSerTyrGlnIleMetThrSerLeuAspValAspTyrVal 609
 DB 1781 ATGGCGTCCACAGAGGAAAGCCATGATGATGAGGAGCTCGATGTCAGCTATGTG 1840
 QY 610 LeuValIlePheGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrp 629
 DB 1841 CTGGTCATTTTGGAGGCTCCTCTGGTATTCCTCTGATGATATCAACAAGTTCTTTGG 1900
 QY 630 MetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluSer 645
 DB 1901 ATGGTCCGATTTGGAGGACACAGATACAGCAACAT-----ATCAGGAGAAAT 1951
 QY 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
 DB 1952 GACTATTATCTCCCAACTGGGAGTTCGGTGGACCGTGAAGTCTTCCAGTGTCTC 2011
 QY 666 AsnCysLeuMetTyrLysLeuSerTyrArgPheGlyGluLeuLysLeuAspTyrArg 685
 DB 2012 AACTGCCATGTCACAGATGTGTACTATCGCTTTGGACAGGTTTACAGAGCCNAG 2071
 QY 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheLeu 705
 DB 2072 CGTCTCCAGGCTTTGACCGTGTCCGAAATGCTGAGATTGGAAATAAGACTTTGAGCTT 2131
 QY 706 ThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLys 725
 DB 2132 GATGCTCTGGAGGAGGTATACCAAGACATTTGGCTGTGCTCAGATATACAGGTAAAG 2191
 RESULT 10
 AB232036
 ID AB232036 standard; DNA; 2256 BP.
 XX
 AC AB232036;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE Candida albicans essential gene SEQ ID NO 6323.
 KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KW signal transduction; DNA replication; cell division; growth;
 KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
 XX
 OS Candida albicans.
 XX
 PN WO200253728-A2.
 XX
 PD 11-JUL-2002.
 XX

PF 26-DEC-2001; 2001WO-US049486.
 XX
 PR 29-DEC-2000; 2000US-0259128P.
 PR 20-FEB-2001; 2001US-00732024.
 PR 22-AUG-2001; 2001US-0314050P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
 XX
 DR WPI: 2002-566694/60.
 DR P-PSDB; ABP73486.
 XX
 PT Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele of
 PT a gene and placing other allele of the gene under conditional expression.
 XX
 PS Claim 37; SEQ ID NO 6323; 167pp + Sequence Listing; English.
 XX
 CC The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office
 XX
 SQ Sequence 2256 BP; 648 A; 336 C; 456 G; 816 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3.35e-212 Length: 2256
 Score: 2175.00 Matches: 434
 Percent Similarity: 70.16% Conservative: 109
 Best Local Similarity: 56.07% Mismatches: 169
 Query Match: 53.78% Indels: 62
 DB: 6 Gaps: 15

US-10-028-384-8 (1-774) x AB232036 (1-2256)

QY 11 LysValAlaGlyTyrSer---SerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeu 29
 DB 34 AAGATTGGCGATAGATGTTGAACCTATTAGACTTTTATTAAGATTATATTATA 93
 QY 30 -----AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIle 45
 DB 94 TCGATAGCAGGTGCAGCTATTTCTTCTCGTTTATTTTCCGTGATTCGATTGAAAGTATT 153
 QY 46 IleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGly 65
 DB 154 ATTATGAAATTCGATCTCTGTTCAATTCGAGCAACCAATATTATGACTCATCTCC 213
 QY 66 TrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleVal 85
 DB 214 TTTTATGAATTTTGAATTCGTTTGTATGATAGATACTTGGTACCATTCGAGAGTCACT 273

Qy 86 GlyGlyThrValThrProGlyLeuMetIleThrSerGlyGlyIleHisThrLeuHis 105
 Db 274 GGTGGTACTTATATCCCGTTTAAAGGTGACCTCAGGTGCCATT---TGG-----CAT 324
 Qy 106 ValLeuAsn-----IleProValHisIleAtrAspIleCysValPheLeuAla 121
 Db 325 ATTTTCGTGATGGTTGGCTTACCCTGTGATATTAGAAATATTGGTTTATTAGCA 384
 Qy 122 ProIlePheSerGlyLeuThrSerIleSerThrThrLeuThrThySclLeuThrPhe 141
 Db 385 CCAGTTTCTCGGATTAACGCAATTGCTACTTATTTTTCGATTAAGCAATGAAGAT 444
 Qy 142 AlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyThrIleSerArg 161
 Db 445 TCTAGTCGAGATTATTGGCAGCTATATTTATGGGATTTGCCCGAGTTATATTTCAAGA 504
 Qy 162 SerValAlaGlySerThrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 181
 Db 505 TCAGTGGCTGGTCTTATGATATAATGAAGCAATTCGCAATTACTTTATTTAATGGCAACATTT 564
 Qy 182 PheLeuTrpValArgSerValLeuThrGlySerValPheThrSerAlaAlaLeu 201
 Db 565 TATTTCTGGATTAATCAATGAATGGGTTCAGTTTCTATGCCCATTTGACAGCATTA 624
 Qy 202 SerThrPheThrMetValSerAlaThrGlyGlyThrValPheIleAlaLeuLeuPro 221
 Db 625 TTCTATTCTATATGTTAGTGGTGGTGGATATGTTTTCATTACCAATTTGATTTCCA 684
 Qy 222 LeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSerTyr 241
 Db 685 TTACACGATTTGCTTGTATTTTCATGGGTCGTTATTAATGCCAACTTTACACTGCTAT 744
 Qy 242 SerThrPheThrLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPheGln 261
 Db 745 ACTACATGTTGGCTTGGGTACTTTGGCATCAATGAGATTCATTCCTGGTGGTTTGA 804
 Qy 262 ProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAlaVal 281
 Db 805 CCAATTAGATCAAAATGATCATATGGCTGCATTTAGGAGTATTTGGATTTGTTACAAATTAGT 864
 Qy 282 AlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgHisLeuPheIle 301
 Db 865 GCTTTTGGTGATTTGTTAAATCAAAAGTTCACCAACCAAAATTTAAATCATTTCTTGATA 924
 Qy 302 ValGly---GlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 Db 925 GTTTCCTATTGACTGTGTGTGGATTAGGTATTT---GGTGGATTTATTTGGATTACAGCA 981
 Qy 321 LeuGlyValValAlaProTrpSerGlyArgPheThrSerLeuTrpAspThrGlyTyrAla 340
 Db 982 ATGGGTTGGATGTCTCTGGACAGGTAGATTTTATTCCTTATGGGATACAAATTTATGCC 1041
 Qy 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 Db 1042 AAGATTCAATTTCCAAATTTGCTTCTGTTTCTGAACATCAACCTACTGCTGGCCAGCA 1101
 Qy 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpThrCysIle 380
 Db 1102 TTTCTTTTCGATAGTATGCTTATTTGGTTATTTCCCGCTGGTATCTATTTATGTTTC 1161
 Qy 381 LysGlnIleAsnAspGluArgValPheValValLeuValLeuTrpAlaIleSerAlaValTrpPhe 400
 Db 1162 CAGAATTGAAGATGAACAGTTTTCATTTATTTACAGTGTATTTGTTCTTATTTT 1221
 Qy 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 Db 1222 GCTGGTGTCTAGTGAAGATTGATGTTGACTTTTTCACCTCCAGTCATTTGTTGCTCGACCA 1281
 Qy 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerIysArgMetGly 440
 Db 1282 ATTGCCCTTAICTAAATTTGTTGATGCTATTG-----AGAAACAGATCCCATTA 1314
 Qy 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThr 460

Db 1315 -----GACATTGTTGATTGTTCACT 1335
 Qy 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThr 480
 Db 1336 -----GAGAAGTTGGAAAGTATGATCATGACCTGTTAGTCAGCAATCCAGAGCTCAACC 1389
 Qy 481 GlyValSerSerAsn-----LeuLysSerIleValIleLeu 492
 Db 1390 AAAAAATCAAGTTCAGAGATTCCCAATTCGCTGATATTGTCACAAAGTTTGGTTTACTG 1449
 Qy 493 AlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAla 512
 Db 1450 ACATTTACATTTTACCTTTTCTACTTTTACATTTGTTACTTGGTAAACATCGAATGCT 1509
 Qy 513 TyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIle 532
 Db 1510 TATTCATCACCATCAGTTGTTTATGCAATCCAGAAAC---CCAGATGGCTCACAACATATC 1566
 Qy 533 LeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAspAspAlaArg 552
 Db 1567 ATGTATGATATAGAGAAGCTTATCTGTTAGAAATGATACACAGAGATGCCAAA 1626
 Qy 553 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeu 572
 Db 1627 GTTATGGCTGGTGGGATTTATGTTTATCAATCGGGGTATGGCTGATGAACACACATT 1686
 Qy 573 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 592
 Db 1687 GTTGTATCAATACATGGAATAACACATATTTGCCACTGTTGGTAGGCAATGCTCTCC 1746
 Qy 593 ThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIle 612
 Db 1747 CCTGAAGATGTCGTATGAAATTTTGACACACACCATGTTGATTATGTTAGTTATA 1806
 Qy 613 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 632
 Db 1807 TTGGNGGTTATGGGTATTTCTGGTATGATATTACAAATTTCTATGATGTTAAGA 1866
 Qy 633 IleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPheThrAspArgGly 652
 Db 1867 ATGTCTGAAGTATCTGGCTGATGAATCAAGAAAGAGACTACTTTACTGCCGAGGA 1926
 Qy 653 GluPheArgValAspAlaGluValAlaProAlaLeuLeuAsnCysLeuMetTyrLysLeu 672
 Db 1927 GAATATAAGTGGATAAAGATGCATCTGCATGATGAAGATTTCTTTGATGTTAAGTTA 1986
 Qy 673 SerTyrTrpArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGlyTyrAspArg 692
 Db 1987 TCGTATCATAGATTTCACCTGAATTTG-----TTGGAGGTAGAGATGGTGTGTAGA 2037
 Qy 693 ThrArgAsnAlaValIle---GlyAsnLysAspPheAspLeuThrTyrLeuGluGluAla 711
 Db 2038 GTTAGAACCAACAATCCAGCCATGAAGTACCGAAATGAATGTTGTTGTAAGAGCC 2097
 Qy 712 TyrThrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHisGluPheAsn 731
 Db 2098 TTCATATCAGAAAAATTTGGATTTGAGAAATTTACAAAGTTAAA----- 2139
 Qy 732 ArgProSerLeuLysThrLysGluArgThrIleProAlaAsnPheIleSerArgLys 751
 Db 2140 -----GATTGGATAATGTTGGTAGATTTTACATCAAGCTACTGCTTTTGAAGATCA 2193
 Qy 752 AsnSerLysArgGlyGlyTyrIleArgAsnArgProVal 765
 Db 2194 TCATCCGCACCTTCCAAA-----AGAAACAGATCCCATTA 2226

RESULT 11
 ID ABT20789
 XX ABT20789 standard; DNA; 2232 BP.
 AC ABT20789;
 XX

DT 16-APR-2003 (first entry)
 DE Aspergillus fumigatus essential gene #3147.
 XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX Aspergillus fumigatus.
 OS WO200286090-A2.
 XX 31-OCT-2002.
 XX 23-APR-2002; 2002WO-US013142.
 XX 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 XX New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX Disclosure; Page; 175pp; English.
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterization, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX
 SQ Sequence 2232 BP; 473 A; 605 C; 545 G; 609 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 7,52e-212 Length: 2232
 Score: 2172.50 Matches: 427
 Percent Similarity: 71.24% Conservative: 108
 Best Local Similarity: 56.86% Mismatches: 171
 Query Match: 53.70% Indels: 45
 DB: 8 Gaps: 11
 US-10-028-384-8 (1-774) x ABT20789 (1-2232)
 QY 18 LeuLeuThrPheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerArgLeuPhe 37

Db 61 CTTCTGGGATTATATCTCTGTGTACCATTTGCAGCAGCAGCATTTGCCAGTCACTCTTC 120
 QY 38 AlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArgAla 57
 Db 121 AGCGTTATCGGCTTCGAGAGTATCATCCAGCAAGATTGACCCGTGGTTCACCTTCGAGCA 180
 QY 58 ThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAla 77
 Db 181 ACAAATACTTAGTACAGATGGTTCTATAGCTTTTGGGATTTGGTTGTATGACGGAACA 240
 QY 78 TrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThrSer 97
 Db 241 TGGCATCTCTCGGACGCTTCACCGGTGGCAGCTTATATCCCGTCTCTCATGGTGCAGC 300
 QY 98 GlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCys 117
 Db 301 GCGGTGATCTACCATATCTTGGGATTCCTTACTATCCCGTGCATATTCGCAACATCTGC 360
 QY 118 ValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLys 137
 Db 361 GTCTTACTGGCGCCAGGATTCTCCGGCTGACTGCATTTGGCAATGTACTTTGCTGACATCC 420
 QY 138 GluLeuTrp---SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
 Db 421 GAGATGTCCTTTCGCCATCTGCAGGCTTCTTTCGACGACGCTTCATGGGAATCGCCCT 480
 QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
 Db 481 GGTATACATCTCCCGATCAGTTGCTGGAAGCTAGCATAACGAGCATTTGCCATCTTTCTG 540
 QY 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSer 196
 Db 541 CTTGTGTTTCACTTCTTCTATGATGATCAGGCTGTCAAAATGGGTCTATCATGTGGGA 600
 QY 187 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 216
 Db 601 GCGGTGACCGCATTATCTACGGTACATGGTGTGGCATGGGTGGTATGTCTTCATT 660
 QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlyArgTyrSerProArg 236
 Db 661 ACGAACCTGATCCCGCTGCGAGTTTGTCTTCTGTGATGGTGTAGATACAGCATCGC 720
 QY 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
 Db 721 ATCTACATTAGCTATACCATGTTGCTGCGCTGCGGACTTTGGCTAGCATGCAGATCC 780
 QY 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
 Db 781 TTGTCGGATTTTGGCTATCCGAAACGAGCAGCACATGTCGCCCTTGGGTGTCTTCGGC 840
 QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
 Db 841 CTGCTTCAGCTTGTGGCTTCGCGAGTTTGTCCGAGCTTCGTTCCAAAGCAAGCAGTTC 900
 QY 297 ArgIleLeu-----PheIleValGlyGlyLeu---LeuValGlyValGly 310
 Db 901 CAGAGATCTTGACCGCCAGCATCTTATCACTTCGGTCTCGTTGCTGGG----- 954
 QY 311 ValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArg 330
 Db 955 -----CTAGTTGTTCTGACTGTGACGGAGTGTATGCTCTCTGGAGCGGCGGA 1002
 QY 331 PheTyrSerLeuTrpAspThrGlyTyrAlaValIleHisIleProIleIleAlaSerVal 350
 Db 1003 TTCTACTCTTGTGGGACACTGGCTATGCCAAATCCAAATCCAAATCCATTCCTGCTCAGTC 1062
 QY 351 SerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuValCys 370
 Db 1063 TCGGAACACACAGCCCGCTGGCCAGCGTCTTCTTCGATCTGAATCTCCCTGATCTGG 1122
 QY 371 AlaPheProValGlyValTrpTyrCysIleGlnIleAsnAspGluArgValPheVal 390

QY	748	IleSerArgIysAsnSerIysArgArgIysGly	758
DB	2176	CTCAAGAAAAGCGGAGTACAAAGGAAAGGGA	2208
<p>RESULT 12</p> <p>ABV24502</p> <p>ID ABV24502 standard; cDNA; 2760 BP.</p> <p>XX AC ABV24502;</p> <p>XX AC</p> <p>XX DT 16-SEP-2002 (first entry)</p> <p>XX DE</p> <p>XX DE Human prostate expression marker cDNA 24493.</p> <p>XX KW Human; prostate cancer; cytostatic; carcinogen;</p> <p>XX KW pharmacogenomic marker; gene; ss.</p> <p>XX OS Homo sapiens.</p> <p>XX PN WO200160960-A2.</p> <p>XX PD 23-AUG-2001.</p> <p>XX PF 20-FEB-2001; 2001WO-US005171.</p> <p>XX PR 17-FEB-2000; 2000US-01831319P.</p> <p>XX PR 16-MAR-2000; 2000US-0189862P.</p> <p>XX PR 25-MAY-2000; 2000US-0207454P.</p> <p>XX PR 09-JUN-2000; 2000US-0211314P.</p> <p>XX PR 18-JUL-2000; 2000US-0219007P.</p> <p>XX PR 13-DEC-2000; 2000US-0255281P.</p> <p>XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.</p> <p>XX PI Schlegel R, Endege WO, Monahan JE;</p> <p>XX DR WPI; 2001-662795/76.</p> <p>XX PT Novel isolated nucleic acid molecule associated</p> <p>XX PT prostate cells and correlating with presence of</p> <p>XX PT for detecting presence of prostate cancer, stages</p> <p>XX PS Claim 1; Page 4632-4633; 11750pp; English.</p> <p>XX CC The invention relates to an isolated nucleic acid</p> <p>XX CC a nucleotide sequence given in Tables 1-9 (ABV00</p> <p>XX CC specification of its complement. (I) is useful</p> <p>XX CC a patient is afflicted with prostate cancer; (b)</p> <p>XX CC progression of prostate cancer in a patient; (c)</p> <p>XX CC of a test compound to inhibit prostate cancer in</p> <p>XX CC the efficacy of a therapy for inhibiting prostate</p> <p>XX CC (e) selecting a composition for inhibiting prostate</p> <p>XX CC (f) assessing the prostate cell carcinogenic pos</p> <p>XX CC determining whether prostate cancer has metastas</p> <p>XX CC assessing the aggressiveness or indolence of pro</p> <p>XX CC ; (I) is also useful as a pharmacodynamic or ph</p> <p>XX SQ Sequence 2760 BP; 688 A; 645 C; 658 G; 769 T; 0</p>			
<p>Alignment Scores:</p> <p>Pred. No.: 2,65e-211</p> <p>Score: 2168.50</p> <p>Percent Similarity: 73.27%</p> <p>Best Local Similarity: 57.76%</p> <p>Query Match: 53.60%</p> <p>DB: 5</p> <p>US-10-028-384-8 (1-774) x ABV24502 (1-2760)</p>			
QY	17	SerLeulleThrPheAlaIleLeuLeulleAlaTrpLeu	
DB	150	ACATTTTAGGCTTCATCTGTCGAAGGCTGCTGTA	

QY 37 phealavalleargphegluSerllelleHieglupheaspProtrpPheasnTyrAr 56
Db 210 TTTTCTCTCTGAGATTGAAGTGTATCTCATGCTTTGATCCGCTACTTTAATATCG 269
QY 56 gAlaThrAlaTyrMetValGlnAenGlyTyrPtyrAsnPhelLeuAenTrpPheaspGluAr 76
Db 270 GACTACCAAGGTTCCCTGGCTGAGGAGGGTTTATAAATTCATTAACCTGGTTGATGACCG 329
QY 76 gAlaTrpTyrProLeuGlyArglleValGlyGlyThrValTyrProGlyLeuMetIleTh 96
Db 330 AGCCCTGGTACCCCTTTGGACCAATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCAC 389
QY 96 rSerGlyGlyIleHisTrpLeuHis-ValLeuAenlleProValHisIleAraAspI 116
Db 390 CTCTGCTCAATCTACCATGATCTCCATCTTTTCCACATCACCATCGACATTCGGAATG 449
QY 116 leCyeValPheLeuAlaProIlePheSerGlyLeuThrSerlleSerThrTyrLeuLeuT 136
Db 450 TCTGTGTCTCTGCGCCCT 509
QY 136 hrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAaSerPheIleAlaIleValP 156
Db 510 CCARAGAGCTCAAGGATCGAGGGCTGGCTTCTGTCTGCTGCCATGATTCGTAGTTC 569
QY 156 roGlyTyrIleSerhrGserValAlaGlySerTyrAspAenGluGlylleAlaIlePheA 176
Db 570 CTGGATATATCTCCGATCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 629
QY 176 laLeuGlnPheThrTyrPheLeuTrpValArgSerValTyrGlySerValPheTrpS 196
Db 630 GCATGCTACTACCTACTACATGATGATCAAGGAGGTAAGACTGGTTCCATCTGTTGG 689
QY 196 erAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheI 216
Db 690 CAGCTAAGTGTGCCCT 749
QY 216 lleAenlleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProA 236
Db 750 TGATCAACTTAATCTCTCCAGCTCTCTGCTGCTGATGCTCACAGGCGCTTCTCTCAC 809
QY 236 rGLeuLeuThrSerTyrThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleP 256
Db 810 GGATCTATGTGGCTTACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 869
QY 256 roPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheV 276
Db 870 CCTTTGGGTTTCCAGCT 929
QY 276 alLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAenGluP 296
Db 930 GTCTCTGCCAGATCCATCCCTTTGTGGATTACCTGGCAGCAATGTGAATCCCAACAAT 989
QY 296 heArgLysLeuPhePhePhePhePhePhePhePhePhePhePhePhePhePhePhe 311
Db 990 TTGAAGTCTTTTCCGAGGCTCATCTCTCTGCTAGCTTTGCTCTCTCTCTCTCTCTCT 1048
QY 311 alPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgP 331
Db 1049 -----GCTCTCTCATGTGACA-----GGAAAAAATATCTCTCCCTGGACGGGCGT 1094
QY 331 heTyrSerLeuTrpAepThrGlyTyrAlaLyslleHisIleProIlelleAlaSerValS 351
Db 1095 TCTACTCGCTGCTGATCCCTCTTATGCTAAGAACACATCCCATCATCTCTCTCTCTCT 1154
QY 351 erGluHisGlnProThrTrpPheSerPhePhePhePhePhePhePhePhePhePhePhe 371
Db 1155 CTGAGCATCAGCCCAACCTGGTCTCTCATATCTTTTGTACCTTCAGCTCTCTCTCTCT 1214
QY 371 laPheProValGlyValTyrPtyrCyslleLysGlnlleAenAspGluArgValPheValV 391
Db 1215 TGTTCCTGCTGCTCTACT 1274

RESULT 13

QY 391 alLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgIleMetLeuThrL 411
Db 1275 TCATGTATGGTGTACAGCATGTACTTTTCTACCTGTAAATGGTGTCTTAATCTAGTGT 1334
QY 411 euThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAenValPheL 431
Db 1335 TGGCACCTGTATGTGATCTCTCTGTCATCTCTGTCATCTGAGTCTCCAGGTCGTCTCCACATCA 1394
QY 431 euGlnGluAenSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAenP 451
Db 1395 TGAAGATCTGGACATAGTCGCTCA-----GACAAAGAGCAAG----- 1420
QY 451 lAlaGluAenSerlleGlyLysThrLeuTyrAspLysAlaGlyLysLeuLysHisA 471
Db 1421 -----GACAAAGAGCAAG----- 1435
QY 471 rGThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerI 489
Db 1436 -----AAGCAACAGATTCACCTACCTATTAAGATGAAGTGGCAAGTG 1481
QY 489 leValIleLeuAlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValT 509
Db 1482 GGATGATCTGCTCATGGCTTCTCTCTCATCTACCTACACCTTTCATTCAACCTGGGTGA 1541
QY 509 hrSerAsnAlaTyrSerSerProSerlleValLeuAlaPheHisAsnSerGlnAspGlyS 529
Db 1542 CCAGTGGCT 1598
QY 529 erArGAsnIleLeuAenAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaA 549
Db 1599 GTAGATCATATTTGATGATCTTCGAGAAGCATATTTATTCGCTTCGTATATAATPACTCCAG 1658
QY 549 spAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnA 569
Db 1659 AGGATGGAGGTCATGCTCTGGTGGATTTATGCTATCAGATTCAGCTATGCGAAACC 1718
QY 569 rGThrTrpLeuValAspAsnAsnTrpAsnAsnSerHisIleAlaLeuValGlyLysA 589
Db 1719 GAACAATTTTAGTGGACAATAACACATGGAATAATACCCATATTTCTCGAGTAGGCGAGG 1778
QY 589 laMetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAenValAspTyrV 609
Db 1779 CAATGGGCTCCACAGAGAAAGCCATGAGATCATGAGGAGCTCGAGGAGCTCGATG 1838
QY 609 alLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuT 629
Db 1839 TGCTGGTCATTTTGGAGGCTCTCACTGGGTATTCCTCTGATGATATCAACAAGTTTCTTT 1898
QY 629 roMetValArglleAla-----GluGlyGluHisProLysAspIleLysGluS 645
Db 1899 GGATGGTCGAGATTTGGAGGAGACACATACAGGCAACAT-----ATCAGAGAGA 1949
QY 645 erAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuL 665
Db 1950 ATGACTATATATCTCAACTGGGAGTTTCCGTGGACCGTGAAGTTTCTCCAGTCTGC 2009
QY 665 euAenCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrA 685
Db 2010 TCAACTGCTCATGACAGATGTTTACTATCGCTTGGACAGGTTTACACAGAGCCA 2069
QY 685 rGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspL 705
Db 2070 AGCGTCTCCAGGCTTTGACCGGTCTCCAAATGTGAGATTGGGAATTAAGACTTTTGAGC 2129
QY 705 euThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArglleTyrArgValI 725
Db 2130 TTGATGCTCTGGAGAGCATATACACAGAACATTTGGCTGGTCCAGATATACAGGTAA 2189
QY 725 ys 725
Db 2190 AG 2191

ADB69900
 ID ADB69900 standard; DNA; 2466 BP.
 AC ADB69900;
 DT 04-DEC-2003 (first entry)
 DE C. neoformans open reading frame SEQ ID NO:2305.
 XX ds; gene; fungicide; gene therapy; infection.
 KW
 XX Cryptococcus neoformans.
 OS
 XX WO2003052076-A2.
 PN
 XX 26-JUN-2003.
 PD
 XX 17-DEC-2002; 2002WO-US040225.
 PF
 XX 17-DEC-2001; 2001US-0341261P.
 PR
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX Zamudio C, Eroskin AM;
 PI
 XX WPI; 2003-533017/50.
 DR
 XX P-ESDB; ADB70261.
 XX
 PT New nucleic acid, useful for preparing a composition for treating an
 FT infection caused by Cryptococcus neoformans.
 TT
 XX Claim 2; SEQ ID NO 2305; 136pp; English.
 PS
 XX The invention relates to a novel purified or isolated Cryptococcus
 CC neoformans nucleic acid molecule comprising a sequence encoding a
 CC polypeptide comprising a sequence not given in the specification. A
 CC polynucleotide of the invention has fungicide activity, and may have a
 CC use in gene therapy. The nucleic acid is useful for preparing a
 CC composition for treating an infection caused by Cryptococcus neoformans.
 CC The present sequence represents a C. neoformans sequence of the
 CC invention. Note: the sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 2466 BP; 508 A; 678 C; 580 G; 700 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 9,678-206 Length: 2466
 Score: 2113.50 Matches: 407
 Percent Similarity: 70.82% Conservative: 122
 Best Local Similarity: 54.48% Mismatches: 175
 Query Match: 52.24% Indels: 43
 DB: 10 Gaps: 10
 US-10-028-384-8 (1-774) x ADB69900 (1-2466)
 QY 4 ThrProLysMetLeuAnSerLysValAlaGlyTyrSerSerLeuLeuThrPheAlaLeu 23
 DB 199 TCTCCGTCGACTATCAACACACG-----GAAAGCTTGTGGCTTCATCATC 246
 QY 24 LeuLeuLeuAlaTrpLeuAlaGlyPheSerArgLeuPheAlaValLeuArgPheGlu 43
 DB 247 CTGCGGTTAATATGCGGTGCTGGATTGGAAGTCGATTCTTTCGGTGCATCAGATTGAA 306
 QY 44 SerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyrMetValGln 63
 DB 307 TCTGTCAATCCACCAATTTGACCCCTGGTTCACACTACCGAGCCCTCGAAGTCTTGTAAAC 366
 QY 64 AsnGlyTrpTyrAsnPheLeuAnTrpPheAspGluArgAlaTrpTyrProLeuGlyArg 83
 DB 367 AAGGGTTCTACGAGTTCTGGAATCTGGTTGACCCCTCGCTTGGTACCTCTCGGCAGA 426
 QY 84 IleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeu 103

DB 427 ACTGTCGGTACACCGCTCTATCTGGCTTGATGTGTACGCTCGTGCATGATTTGGCATGCT 486
 QY 104 LeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeuAlaProIle 123
 DB 487 CTTCCGGCAATCAATATGCCCGTGGACATTCGCAATCTGTGTCTCTCTTGCACCTGGA 546
 QY 124 PheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuLeuTrpSerAlaGly 143
 DB 547 TTTTCTGGATTGACTGCTCGCGGCACTATCTTTTCCACCTGAAATGTCTACCATCA 606
 QY 144 AlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSerArgSerVal 163
 DB 607 GCTGGTCTATTGGCGCGCTTTCATTGGCATTTACCCGATACATCTCTCGATCTGTC 666
 QY 164 AlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeu 183
 DB 667 GCGGTTCTTATGACACGAAGCAATGCCATCTCTCTTGATGAGCTCTCTTACTCT 726
 QY 184 TrpValArgSerValIleThrGlySerValPheTrpSerAlaAlaAlaLeuSerTyr 203
 DB 727 TGGATTAAAGCGCTCAAAACCGGTAGCTCAATTTGGGGTATGATCACTGCTTGTCTAC 786
 QY 204 PheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHis 223
 DB 787 GGGTGGATGGTTCGTCATGGGGTGGTTCAGTTTTCATCAACCAATGATTCATTCGAC 846
 QY 224 ValPheValLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThr 243
 DB 847 GCCTTTGTCTCATTTGTCATGGCAGGTTCACAAACCGCTTTATACCGCTTACTCTTCC 906
 QY 244 PheTyrIleLeuGlyLeuLeuPheSerMetClnIleProPheValGlyPheGlnProIle 263
 DB 907 TGGTATGTCATTTGGAATCATCGCTTCATGAGTCCCTTTTGGAGTTTCTCCCATC 966
 QY 264 ArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThr 283
 DB 967 CGAACCTCTGAGCATGCGCGCTTGGTGTCTTTCGGTCTGTGACAGCTGATCGGATTC 1026
 QY 284 LeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArg-----LysLeuPhe 300
 DB 1027 GTCCGAAGTCGTCGACGACGCTCGCTCCGCAAGCAATTCAGTCTCTTCAAAGTTT 1086
 QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 DB 1087 GTCGTCGGCGTATTCTGCTCAGT-----TTTCTGCGCTCGTCACTTTGACTTTC 1137
 QY 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 DB 1138 TCTGGATGGATGCGCCCTTCGCTGGAAGATTTTATTCTTTGGGATACTGCTATGCG 1197
 QY 341 LysIleHisIleProIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 DB 1198 AAGGTCCACATGCCCATTTATTCCTCGCTCCGACACAGCCCGCTTGGCCTCA 1257
 QY 361 PhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrPheCysIle 380
 DB 1258 TTTCTACTTGTACCTCGAAATGCTTATCTTTTCCCTCGGNGTCITCTGGTGTTC 1317
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 DB 1318 AAGGAGCTTCGGATGAGCAGATCTTCATCATCATTTATCCGCTTTCAGCTGCTATTT 1377
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 DB 1378 GCGGTGTTCATGGTTCGACTTATGCTTGTATCATCAGCTTGTGTCTTCTTCTCCGCC 1437
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 DB 1438 ATTGGGTTCTCCAAACTTCTCGAGCGCTATATTGAC----- 1473
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGlyLysThr 460

AC ABT20191;
 XX 16-APR-2003 (first entry)
 XX Aspergillus fumigatus essential gene #2549.
 DE Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX Aspergillus fumigatus.
 OS WO200286090-A2.
 XX 31-OCT-2002.
 XX 23-APR-2002; 2002WO-US013142.
 XX 23-APR-2001; 2001US-0285697P.
 XX 27-APR-2001; 2001US-0287066P.
 XX 05-JUN-2001; 2001US-0295890P.
 XX 09-JUL-2001; 2001US-0303899P.
 XX 31-AUG-2001; 2001US-0316362P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Jiang B, Tishkoff D, Zamudio C, Broshkin AM, Hu W, Lemieux SM;
 PI WPI; 2003-093124/08.
 XX New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX Disclosure; Page; 175pp; English.
 XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterization, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organisms invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX SQ Sequence 2603 BP; 565 A; 698 C; 620 G; 720 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.64e-195	Length:	2603
Score:	2014.00	Matches:	429
Percent Similarity:	62.17%	Conservative:	110
Best Local Similarity:	49.48%	Mismatches:	167
Query Match:	49.78%	Indels:	162
DB:	8	Gaps:	17

US-10-028-384-8 (1-774) x ABT20191 (1-2603)

QY	2	AsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThrPhe	21
DB	139	TCGAGGAATCAAAAGTTGCTAAATTTGTTAACTA-TCAATCAAAATGACTGATCACTTTTC	197
QY	22	AlaIleLeuLeuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArg	41
DB	198	-----TTTCTACCAGGC-----	209
QY	42	PheGluSerIleIleHisGlu-----	48
DB	210	TTCCAGAGATATCATCCACGAATGTAAGTATAGTCAATTGTATACCTACGGCCTTCGGCC	269
QY	49	-----PheAspProTrp	52
DB	270	TTTAAACGACCTCGCTTGGACGAATTGATCTACTGCTCGTGATATAGTGTACCCGCTG	329
QY	53	PheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrp	72
DB	330	TTCAACTTCGAGCAACAAAATACTTAGTACAGATGGTTTCTATAGCTTTTGGGATTGG	389
QY	73	PheAspGlu-----	75
DB	350	TTTGTATGACCGTATGTGTTTCCCGCACCAAGAGTGCATGAAGCAACAGGTCACTCT	449
QY	76	-----Arg-AlaTrpTyrProLeuGlyAr	83
DB	450	CCTGGAGTGTCTTAGCTAATCGCGTCCCAACATAGGAACATGCGCTCTCTGGGACG	509
QY	83	GileValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLe	103
DB	510	TGTCACCGGTGGCAGCTTATCCCGTCTCATGGTGACGAGGGCGTGATCTACCATAT	569
QY	103	uLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeuAlaProI	123
DB	570	CTTCGATTCCTTACTATCCCTCGCATATCCCAACATCTCGCTCTCTACTGGCGCAGG	629
QY	123	ePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp---SerAl	142
DB	630	ATTCTCCGGCTGACTGCATTCGATGCTGCTGATCGAGATGTCTCTCTCGGC	689
QY	142	aGlyValGlyLeuPheAlaLaserPheIleAlaIleValProGlyTyrIleSerArgSe	162
DB	690	ATCTGCAAGGCTTCTTCGACAGCTTTCATGGGAATCGCCCTGGTATACCTCCGATC	749
QY	162	rValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPh	182
DB	750	AGTTGCTGGAAGCTACATACAGAGCATTCGCATCTTCTCTGTGTTCACATCTT	809
QY	182	eLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAlaLeuSe	202
DB	810	TCTATGGATCAAGCTCTCAAAATGGGTCTATCATGTGGGAGCGCTGACCGCACTAT	869
QY	202	rTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleAsnLeuIleProLe	222
DB	870	CTACGGTACATGGTGTGCGCATGGGTGGGTATGTCTTCATTTACGAACCTGATCCCT	929
QY	222	uHisValPheValLeuLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSerTyrSe	242
DB	930	GCAGCTTTTGTCTCTGTGCATGGGTAGATACAGCATTCGCATACATAGCTATAC	989
QY	242	rThrPheTyrIleLeuGlyLeuPheSerMetGlnIleProPheValGlyPheGlnPr	262
DB	990	CACATGGTATCGCTGGGACTTTGGCTAGCATGCAGATTCCTCTCGTGGATTTTGGC	1049
QY	262	oileArgThrSerGluHisMetAlaAlaLeu-----	272
DB	1050	TATCCGAACAGCGACCATGTCGCGCTTGGGTAGTCTCCTTTCGCAAGTTCAAAT	1109
QY	273	-----GlyValPheValLeuLeuMetAlaValAlaThrL	284
DB	1110	GGCGTTTCAGAGCTGACATTTTCAGGTGTCTTCGGCTGTCTTACGCTTGTGGCCTTCG	1169

284	euArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeu	229
1170	CCGAGTTTGTCCGCGAGCTTCGTTCCCAAGCAAGCAGTTCAGAGACTTCGACGCCCATGA	1229
300	--PheIleValGlyGlyLeu--LeuValGlyValGlyValPheValAlaValValVal	318
1230	TCCTTCATCACCTTCGGTCTCGTTTGGTGGG	1271
318	euThrMetLeuGlyValValAlaProThrPheSerGlyArgPheTyrSerLeuThrPheThrG	338
1272	TCAGTGTACGGGAGTATCGCTCCTTGAGCGGGCGGATCTTACTCTTGTGGACACTG	1331
338	IYTrAlaValLeuIleHisIleProIleIleAlaSerValSerGluHisGlnPheProThrThr	358
1332	GCTATGCCAATAATCCATATCCCATTCCTAGCTCGCTCGGAACACCGCCACCGCTT	1391
358	trpPheSerPhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpT	378
1392	GCCACAGCGTCTCTTCGATCTGAATCTCTGATCTGGCTTTTCCGCGAGGTGCTTACA	1451
378	yrCysIleLeuGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlav	398
1452	TGTGCTTCGGTGACCTCAAGACGACGATGTCTTGTCATTAATCTACTCGTCTCGCGA	1511
398	alTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrLeuThrProValValCysMetL	418
1512	GCTACTTTCGCGGGTGTATCGTCCGACTAATGCTGACCTTGACCCCTATTGTGTGTGTG	1571
418	euaIaGlyValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysA	438
1572	CGCGTGTCTGGCGCTGTGTGTCATCTCGACACCTATATGCGCACT	1618
438	rgMetGlyThrAlaIleSerAlaAlaThrGluValAspGluAlaAspSerIleGluL	458
1619	-----ACCTCCGACACGAGCTGTGAAGCAACGAATGAAGACTCGTTCAC	1670
458	ysLysThrLeuTyrAspLysAlaGlyLeuLysHisArgThrLysHisAspAlaGlnG	478
1671	MAACT	1694
478	InAspThrGlyValSerSerAsnLeuLysSerIleValIleLeuAla--ValLeuMetL	497
1695	CCAGTGTGGAAATCACCTCCCATGTTCTTAAGATTATAGTACGGCGTCTGTGTCTCT	1754
497	euLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProS	517
1755	ACCTGCTCTGTTTGTGGCGCACTGCACCTGGGTACATCGAATGCATCTCTCTCTCT	1814
517	erIleValLeuAlaPheHisAsnSerGluAspGlySerArgAsnIleLeuAspAspPheA	537
1815	CCGTGTGCTTGCT--AGTCGGATGCCTGACGGAAGCCATACATCATTTGACGATTATC	1871
537	rgGluAlaTyrTyrTrpLeuSerGluAsnThrAlaAspAspAlaArgValMetSerTrpT	557
1872	GTGAGGCTTACTATCGTCTTCGTCAGAACTCCTCAGAACGCCAAAATCATGTGTCATGGT	1931
557	rpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnt	577
1932	GCGATTATGGTATCAATCGGTGGCATGGCGGACGCGCCAACTTGGTTGACACACACA	1991
577	hrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerSerThrGluGluLysS	597
1992	CCTGGAAACAACCCATTTGCTACCGTTGGTAAGCGCATGAGCTCACGGGAGGAAGTCA	2051
597	erTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIlePheGlyValI	617
2052	GCTACCCCATCTCCGCCACGATGATCGATTACGTGCTGGTGGTGTTCGGTGTCTGC	2111
617	leGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyG	637
2112	TAGGTATCTCGCGCATGACATTAAACAAATCTTATGATGTCGTATGCGCAAGGTA	2171
637	IuHisProLysAspIleLysGluSerAspTyrPheThrAspArgGlyGlyLupheArgValA	657

Search completed: December 14, 2004, 20:05:38
Job time : 1020.03 secs

QY 35 ArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPheAspProThrPheAsn 54
 Db 94 CGITTTGTTGCGATCATCAAAATTGAGTCTATTATCCCAATTCGACCCCTGGTTCAAT 153
 QY 55 TyrArgAlaThrAlaTyrMetValGlnAenGlyTyrTyrAsnPheLeuAsnTrpPheAsp 74
 Db 154 TATAGGCTACCAAAATATCTCGTCAACAATTCGTTTACAAAGTTTTCGAAGCTGGTTGAC 213
 QY 75 GluArgAlaTrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMet 94
 Db 214 GACCGTACCTGGTACCCCTCGGAAGGGTACTCGAGGAGCTTTATATCTCGTTGATG 273
 QY 95 IleThrSerGlyGlyIle---HistripLeuLeuHisValLeuAsnIleProValHisIle 113
 Db 274 ACGACTAGTCGTTCACTCGGCACGCCCTGGCAACTGGTTGGGCTTGCCTTGCACATC 333
 QY 114 ArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyr 133
 Db 334 AGAACGTTTGTGTGCTATTGCGCCACTATTCTCGGGTCCACCGCTGGCGACTTAC 393
 QY 134 LeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAla 153
 Db 394 GAATTTACGAAAGAGATTAAAGATGCGCGCTGGGCTTTGGCTGCTGGTTTATAGCC 453
 QY 154 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAsnAsnGluGlyIleAla 173
 Db 454 ATTGCCCCGGTTATATATCTAGATCAGTGGCGGGTCTCTAGATPANTAGGCCATTCGCC 513
 QY 174 IlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValTyrGlySerVal 193
 Db 514 ATTACACTATTAAATGGTCACTTTTCATGTTTGGATTAAAGCCCAAAAGACTGGCTCATC 573
 QY 194 PheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyTyr 213
 Db 574 ATGCACACACGTTGAGCTTTTATTCTACTTCTACATGGTTCGGCTGGGTGGATAC 633
 QY 214 ValPheIleLeuAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyr 233
 Db 634 GTGTTTCATCCCAACTGATCCCATGCTTTTCTGCTGATTTTGTATGGCGAGATAT 693
 QY 234 SerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMet 253
 Db 694 TCGTCCAACTGATTTCTGCTACACACTTGTGACCTATTGGAACTTTGTCATCCATG 753
 QY 254 GlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGly 273
 Db 754 CAGATCCCATTTGTCGGTTTCTTACCTATCAGTCTTAACGACCATGCGCGCATTTGGGT 813
 QY 274 ValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArg 293
 Db 814 GTTTTCGGTTTGAATCGATTGTCGCTTCGGTGACTTCGTGAAGGGCCAATCAGACA 873
 QY 294 AsnGluPheArgLysLeuPheIleValGlyGlyLeuLeuValGlyValGlyValPheVal 313
 Db 874 GCTAAGTTTAAAGTCATCATGATGGTTCTCTGTTTTCATCTTGTCTGCTGGTGGTGC 933
 QY 314 AlaValValValLeuLeuMetLeuGlyValValAlaProTrpSerGlyArgPheTyrSer 333
 Db 934 GGACTTTCTGCTTGAATATATGGGGTGTATGCGCTTGCCTTGCCTGAGATTATTCG 993
 QY 334 LeuTrpAspThrGlyTyrAlaLysIleHisIleProIleAlaSerValSerGluHis 353
 Db 994 TTATGGGATACCAACTACGAAAGAGATCCCATTCCTATCATTCCTCGCTTTCGAAAT 1053
 QY 354 GlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuValCysAlaPhePro 373
 Db 1054 CAACCCGTTTCGGTGGCGGCTTCTTCTTTGTATACCCACTTTTGTATCTGGCTATTCGC 1113
 QY 374 ValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValValLeuTyr 393
 Db 1114 GCGCGTGTATCTCTACTTATTCCTCGACTTGAAGACGACGACGCTTTTGTCTATCGCTTAC 1173
 QY 394 AlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrPro 413

1174 TCCGTTCTGTGTTGCTACTTTGCCGGTGTATGTTAGATTGATGTTGACTTTGACACCA 1233
 QY 414 ValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeuGlnGlu 433
 Db 1234 GTCATCTGTGTCCGCGCGTGCATTCGCAAGATATTTCACATCTACTCTGGATTTC 1293
 QY 434 AspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGluAlaGlu 453
 Db 1294 AAGACAAGTGACCGC----- 1308
 QY 454 AspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArgThrLys 473
 Db 1309 -----AAATACGCCATCAAA 1323
 QY 474 HisAspAlaGlnAspThrGlyValSerSerAsnLeuLysSerIleValIleLeuAla 493
 Db 1324 CCTGCGGCACCTACTG-----GCCAAATTGATTGTTCCGGATCA 1362
 QY 494 ValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyr 513
 Db 1363 TTCATCTTTTATTTGATCTTTTCGTTCTTCATCTACTTGGGTAAACAAGACTGCGATAC 1422
 QY 514 SerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIleLeu 533
 Db 1423 TCTTCTCTCTCTGTTGTTTGCCA---TCACAAACCCACAGATCGTAAATTGGCGTTGATC 1479
 QY 534 AspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgVal 553
 Db 1480 GACGACTTCAGGAGGCGTACTATTGTTAAGAAATGAACCTCTGTGAGGACAGTAAGGTT 1539
 QY 554 MetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuVal 573
 Db 1540 GCAGCGTGGTGGATTACGGTTTACCAATTTGGTGGCATGCGACAGACCACTTTAGTC 1599
 QY 574 AspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerThr 593
 Db 1600 GATAACACACGCGGAACAATACATCAGATCGCATCGCTTGGTAAAGCCATGGCTCCCT 1659
 QY 594 GluGluLysSerTyrGluIleMetSerLeuAspValAspTyrValLeuValIlePhe 613
 Db 1660 GAAGAGAAATCTACGAAATCTTAAAGAGCATGATGTCGATTATGCTTGTGTCATCTTT 1719
 QY 614 GlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIle 633
 Db 1720 GTGTCTTAATTTGGGTTTGGTGGTATGACATCAACAAATCTTGTGGATGATCAGAAAT 1779
 QY 634 AlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPheThrAspArgGlyGlu 653
 Db 1780 ACCGAGGGAATCTGCCAGAGAGATAAAGAGCGTGATTTCTATACCGCAGAGGAGAA 1839
 QY 654 PheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSer 673
 Db 1840 TACAGAGTAGATGCAAGGGCTTCTGAGACCATGAGGAACCTCGCTACTTTCACAGATGTC 1899
 QY 674 TyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGlyTyrAspArgThr 693
 Db 1900 TACAAAGATTTCCCACAATTA-----TTCAATGCTGCCCAAGCCACTGACAGATG 1950
 QY 694 ArgAsnAlaValIleGlyAsnLysAspPhe---AspLeuThrTyrLeuGluAlaTyr 712
 Db 1951 CGTCAACAAATGATCACCATTAGAGCTGCCCATCTAGACTACTTTCGACGAAGTTTTT 2010
 QY 713 ThrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHisGluPheAsnArg 732
 Db 2011 ACTTCCGAAACTGGATGTTTAGAATATATCAATTGAAGAGGAT----- 2055
 QY 733 ProSerLeuLysThrLysGluArgThrIleProAlaAsnPheIleSerArgLysAsn 752
 Db 2056 -----GATGCCCAAGGTAGAACCTTTTGAGGACGCTTGGTGAAGTTACACGCTTCT 2106
 QY 753 SerLysArgArgLysGlyTyrIleArg 761

Db 2107 ACGAAACACGAGGTCCATAAGAGA 2133

RESULT 2

US-09-270-767-12331

; Sequence 12331, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-034

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12331

; LENGTH: 900

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-12331

Alignment Scores:

Pred. No.: 1.66e-160 Length: 900

Score: 1435.00 Matches: 273

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 35.47% Indels: 0

DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-12331 (1-900)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20

Db 82 ATGAATCGGACGCCGAAGATGCTGAACAGCAAGGTGGCTGGCTACACGACCTTAATCAAC 141

QY 21 PheAlaIleLeuLeuIleAlaTrpPheuladglyPheSerSerArgLeuPheAlaValIle 40

Db 142 TTCGCCATCCTCTAATGCCGTGGCTGGCGGATTTTCTCTCGCCCTTTTCGGCGTCATC 201

QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60

Db 202 CGTTTCGAGTCGATTATCCATGAGTTTCATCGTGGTTCACATACCGGGCCACCGCCTAC 261

QY 61 MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80

Db 262 ATGGTCGAAATGGTGGTGACAACTTCTCAACTGGTTTCGACGAGCGCGCATGGTATCCG 321

QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100

Db 322 CTCGGCAGGATTGTGGCGGTACCGCTCTATCCGGCGCTGATGATTACGTCGGCGGAATC 381

QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120

Db 382 CARTGGCTGCTGACGTACTCAACATACCGGTCCATATCGTGACATCTCGGTGTTCCGTG 441

QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140

Db 442 GGGCCGATCTTCAGTGGCGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 501

QY 141 SerAlaGlyValGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160

Db 502 TCGCGGGCGCGCGCTCTTCGCCGCCAGCTTCATCGCCATCTGGTGGCTGCATACATCAGT 561

QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180

Db 562 AGGTCGGTGGCTGGATCGTACGATAACGAGGGCATTTGCCATATTTCGCCCTTCGAGTTCCAC 621

QY 181 TyrPheLeuTrpValArgSerValIlysthrGlySerValPheTrpSerAlaAlaAlaAla 200

Db 622 TACTTCCTGGGTGGCTTCAGTGAAGACTGATCCGTTTCTGGTGGCGCGGAGCGCT 681

QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIle 220

Db 682 TTGTCCTACTTCTACATGGTTCGCCCTGGGGTGGCTACGTGTTTCATCATCAACCTGATA 741

QY 388 lPheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMe 408
 DB 457 TGAGTAGCTTAATTTGTACATAGAGCAGTTGAGCAGGCGGACACCTTCGGCATCTA 516
 QY 408 tLeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAs 428
 DB 517 CCCTGGAATTCACGCGGCTGGTAAGTAATCGCTTTCCCTTAATGCTTCCTGGGATGCTCTC 576
 QY 428 pValPheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGl 448
 DB 577 CCTC----- 580
 QY 448 uValAspGluAlaGluAspSerIleGluLysThrLeuTyrAspLysAlaGlyLysLe 469
 DB 581 -----GGCAATTCGACCATCCACAGCACTGTGTATCATCGCGGAAT 627
 QY 468 uLysHisArgThrLysHisAspAlaGlnAspThrGlyValSerAsnLeuLysSe 488
 DB 628 AGCGATCACACGCCCAAGATCACAA-----AACGTAGTCCA 666
 QY 488 rIleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpVa 508
 DB 667 CGTCAAGAGATGTCAATAATTCGTAGGACTTCTCTCGGTTGAGACATTCG-----CT 720
 QY 508 lThrSerAsnAlaTyrSerSer-----ProSerIleValLeuAlaPheHisAsnSerGl 526
 DB 721 TGCCAACCGCGCTATGTGACTATTGTTCCAGTATTATTATCCACTAGCGTCTGTGT 780
 QY 526 nAspGlySerArgAsnIleLeuAspPheArgGluAlaTyrTyrTrpLeuSerGlnAs 546
 DB 781 TTGCATTCCTCCGCAACATTTTACAGCATTTTCAGAGAGGCTTACTACTCGCTTTTCGCAAA 840
 QY 546 nThrAlaAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMe 566
 DB 841 CACTGCCGATGATGCTCGCTTATGCTTGGTGGGATACGGATACCGATAGCGGAT 900
 QY 566 tAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuVa 586
 DB 901 GGCACACAGAACACGCTAGTGGATAATATACGTGGAAACATAGTACATAGCGCTGT 960
 QY 586 lGlyLysAlaMetSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspVa 606
 DB 961 TGCCAAAGCAATGCTTCACCGCAGGAGAGTCTTACGAAATATGACATCTCTGAGT 1020
 QY 606 lAspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLy 626
 DB 1021 GGACTAGCTTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGCGGATGATATACAA 1080
 QY 626 sPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAs 646
 DB 1081 GTTCTCTGGATGCTCGGATTCCTGAGGAGAGCATCCACAGGACATAGGAAGCGA 1140
 QY 646 pTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAs 666
 DB 1141 TTACTTTACGACCGCGGTGAATTCAGGGTAGATCCGAGGTGCTCGGCGCTCTCAA 1200
 QY 666 nCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGl 686
 DB 1201 CTGCTTATGTACAAATTAAGCTACTACAGATTCCGGGAAATGAAGTTGGACTACAGAG 1260
 QY 686 yProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuTh 706
 DB 1261 TCCATCTGGATATGTCGACAGTAAACCGGTCATTGGGAATAGGACTTCGATCTGAC 1320
 QY 706 rTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLysLy 726
 DB 1321 CTACTCTGAGAGCGCTTACACCAACAGACACTGGCTTGTTCGACTATAGGTTGAGAA 1380
 QY 726 sProHisGluPheAsnArgProSerLysThrLysGluArgThrIleProProAlaAs 746
 DB 1381 GCCCATGAGTTCATAGACCATCTCTGAAGACCAAGAGAGCAATCTCTCCAGCAAA 1440
 QY 746 nPheIleSerArgLysAsnSerLysArgLysGlyTyrIleArgAsnArgProValVa 766

DB 1441 CTTCAATTCGAGAAAGAACTCTAAGCTCGCAAGGGCTACATACGAACCGACCGTTGT 1500
 QY 766 lVallysglyLysArgThrIleuLys 774
 DB 1501 TGTTAAGGAAACGAACCTTGA 1525
 RESULT 4
 US-09-270-767-15440
 : Sequence 15440, Application US/09270767
 : Patent No. 6703491
 : GENERAL INFORMATION:
 : APPLICANT: Homburger et al.
 : TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 : FILE REFERENCE: File Reference: 7326-094
 : CURRENT APPLICATION NUMBER: US/09/270,767
 : CURRENT FILING DATE: 1999-03-17
 : NUMBER OF SEQ ID NOS: 62517
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 15440
 : LENGTH: 1660
 : TYPE: DNA
 : ORGANISM: *Drosophila melanogaster*
 US-09-270-767-15440
 Alignment Scores:
 Pred. No.: 4, 4e-147 Length: 1660
 Score: 1326.00 Matches: 305
 Percent Similarity: 62.48% Conservative: 38
 Best Local Similarity: 55.56% Mismatches: 121
 Query Match: 32.77% Indels: 86
 DB: 4 Gaps: 11
 US-10-028-384-8 (1-774) x US-09-270-767-15440 (1-1660)
 QY 245 TyrIleLeuGlyLeuLeuPheSerMetClnIleProPheValGlyPheGlnProIleArg 264
 DB 80 TATGTGCTATGTGTAAGTGTTTTAAACCAAGTTATATATATATGTTGTAAGATCAATACAA 139
 QY 265 ThrSerGluHisMetAlaLeuGlyValPhe----- 275
 DB 140 TTA-----TTTTTTTTTTCGGTTTATTTCAAGTTTCGTTTCCCTTAAACAACAAC 190
 QY 276 -----ValLeuMetAlaValAlaThrIleuArgHisLeuGlnSerValLeuSerArg 293
 DB 191 CGGTGGGTTTCTGTATGTAGACCTTTCGACGCTTCA-----GTCTTTCTCGA 238
 QY 294 AsnGlu-PheArg-----LysLeuPheIleValGlyLeuLeuValGl 308
 DB 239 GATGAAGTTTGTGAGGAATCGTTCTCTCTCTCTTCAGTGTCTATTGAATC 298
 QY 308 yValGlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProIrpse 328
 DB 299 ATGCGCTTCTTCACATATAGATGCGAACAGCCAGTGTCTGTGTGTA----- 349
 QY 328 rGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleAl 348
 DB 350 -----GGCTCTCTCCAGGTAGTGCAGATCAAAGTCTTATTTCCCAATCACCG 396
 QY 348 aserValSerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLe 368
 DB 397 C-----GT 399
 QY 368 uValCysAlaPheProValGlyValTyrCysIleLysGlnIleAsnAspGluArgVa 388
 DB 400 TAGGTGTGCGATC---ATATCCAGATGGGCTCTGTAGTCCAACTCAATTCGCCAATC 456
 QY 388 lPheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMe 408
 DB 457 TGTAGTAGCTTAATTTGTACATAGGCGATTGACAGGCGCGGACGACCTTCGGCATCTA 516
 QY 408 tLeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAs 428


```

QY 475 pAlaGlnGlnAspThrGlyValSerSerAsnLeuIysSerIleValIleLeuAlaVal 495
Db 1037 TCACCAAAAGTATGCCACGTCAGAGATGTCATATTTTCGTAGGACTTCTCTCGG-TT 979
QY 495 uMetLeuLeuMetMetPheAlaValHisCysThrTTPValThrSerAsnAlaTyrSerSe 515
Db 978 GAAGACATTCCTTGCACACGCGTATGTGACTATTGTCACGATATTATATCCACT 919
QY 515 rProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAg 535
Db 918 AGCGTCGTTCTGTTGTCATTCCTCGTATCT-----GGTATCCCGTAATCCACCAAGAC 865
QY 535 pPheArgGluAlaTyrTyrTrpLeuSer-GlnAsnThrAlaAspAlaArgValMetS 555
Db 864 ATACCGCGACATCATCGGAGTGTCTCGAAGGCCA-----GTAGTAA 820
QY 555 eTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspA 575
Db 819 GCCTCTCTGAATCGTCTAAATGTTGCGGGAATGGCAAAACAGACGCTAGTGGATA 760
QY 575 snAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluG 595
Db 759 ATATACGTGGAAACATAGTACATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGG 700
QY 595 luLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIlePheGlyG 615
Db 699 AGAAGTCTCTACGAAATATACATCTCTTGACGTGACTACGTTTGTGTGATCTTTGGCG 640
QY 615 lyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaG 635
Db 639 GTGTGATCGCTATTCCGGCGATGATATCAACAGTTCCTGTGGATGTCGGAATGGCG 580
QY 635 luGlyGluHisProLysAspIleLysGluSerAspTyrPheThrAspArgGlyGluPheA 655
Db 579 AGGAGAGACATCCCAAGGACATTAAGGAAAGCGATTACTTTTACCACCGCGGTGAATCA 520
QY 655 rgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrT 675
Db 519 GGGTAGATCGGAAGGTGCTCCGCCCTGCTCAACTGCTTATGACAAATTAAGTACT 460
QY 675 yrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgA 695
Db 459 ACAGATTCCGGGAATTGAAGTTGGACTACAGAGCCCACTCTGGATATGATCGACACGTA 400
QY 695 snAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrG 715
Db 399 ACGCGCTCATTTGGGAATAAGGACTTTTGATCTGACCTACCTGAGGAGGCTTACACACAG 340
QY 715 luHisTrpLeuValArgIleTyrArgValLysLysProHisGluPheAsnArgProSerL 735
Db 339 AACACTGGCTGTGTGCGATCTATAGTGTGAAGAGCGCGATGATTCATAGACCATCAC 280
QY 735 euLysThrLysGluArgThrIleProAlaAsnPheIleSerArgLysAsnSerLysA 755
Db 279 TGAAGACCAAGGAGAGACGATTCCTCCAGCAAACTTCATCCGAGAAGAACTCGAAGC 220
QY 755 rgArgLysGlyTyrIleArgAsnArgProValValLysGlyLysArgThrLys 774
Db 219 GTCGCAAGGGCTACATACGAAACCGACCGGTTGTTGTTAAGGAAACGAACTTGAA 161

```

RESULT 6

```

US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440

```

```

; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Alignment Scores:
Pred. No.: 1,24e-121 Length: 1660
Score: 1112.00 Matches: 274
Percent Similarity: 52.73% Conservative: 32
Best Local Similarity: 58.53% Mismatches: 136
Query Match: 27.48% Indels: 81
DB: 4 Gaps: 9

US-10-028-384-8 (1-774) x US-09-270-767-15440 (1-1660)

QY 272 LeuGlyValPhe-----ValLeuLeuMetAla 280
Db 1534 TTGGGTTTATTTCAGGTCCTTTTCCTTTAACAAACCGGTCGGTTTCGTATGTAGCC 1475
QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGlu-PheArg-----297
Db 1474 CTTGGAGCGCTAGA-----GTTCTTTCTCAATGAAGTTTCTCGAGGAAT 1427
QY 298 -----LysLeuPheIleValGlyLeuValGlyValGlyValPheValAlaVala 315
Db 1426 CTTTCTCTCTTGGTTCAGTGATGGTCTATTGAACCTATCGCGCTTCTTACCCTATA 1367
QY 315 lValValLeuThrMetLeuGlyValAlaProTrpSerGlyArgPheTyrSerLeuTr 335
Db 1366 GATGCGAACCAAGCAGCTGTTCTGTGGTGA-----GGCCTCTCT 1329
QY 335 pAspThrGlyTyrAlaLysIleHisIlePheIleAlaSerValSerGluHisGlnTr 355
Db 1328 CCAGTAGTCAGATCGAAGTCCTTATCCCAATGACGGC-----1289
QY 355 oThrTrpPheSerPhePheAspLeuHisIleLeuValCysAlaPheProValG 375
Db 1288 -----GTTACGTCGCGATC---ATATC 1269
QY 375 yValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValValLeuTyrAlaTr 395
Db 1268 CAGATGGACCTCTGTAGTCCCACTCAATCCCGAATCTGTAGTACTTAATTGTACA 1209
QY 395 eSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuTyrProValVa 415
Db 1208 TAAGCAGTTGAGCAGCGCGAGCCTTCGCGATCTACCTCAATTACCCGCGTCGG 1149
QY 415 lCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSe 435
Db 1148 TAAAGTAAATCGCTTTCTTAATGTCCTTGGATGCTCTCCCTC-----1106
QY 435 rSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSe 455
Db 1105 -----AGCAATTC 1098
QY 455 rIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAs 475
Db 1097 GGACCATCCACAGAACCTGTGATATCATCCCGCAATAGACCGATCACACCCCAAGA 1038
QY 475 pAlaGlnAspThrGlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLe 495
Db 1037 TCACCAAAACGTCAGTCCACGCTCAAGAGATGTCATATTTTCGTAGGACTTCTCTCGG-TT 979
QY 495 uMetLeuLeuMetMetPheAlaValHisCysThrTTPValThrSerAsnAlaTyrSerSe 515
Db 978 GAAGACATTCCTTGCACACGCGTATGTGACTATTGTCACGATATTATATCCACT 919
QY 515 rProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAs 535
Db 918 AGCGTCGTTCTGTTGTCATTCCTCGTATCT-----GGTATCCCGTAATCCACCAAGAC 865
QY 535 pPheArgGluAlaTyrTyrTrpLeuSer-GlnAsnThrAlaAspAlaArgValMetS 555
Db 864 ATACCGCGACATCATCGGAGTGTCTCGAAGGCCA-----GTAGTAA 820

```

Db 864 ATAACGCGAGCATCATCGCGAGTGTCTTCGAAAGCCA-----GTAGTAA 820
 QY 555 exTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspA 575
 Db 819 GCCTCTCTGAAATCGTCTAAATGTTGGGAATGGCAACAGACGCGTAGTGGATA 760
 QY 575 snAsnThrTrpAsnAsnSerHisIleAlaIleValGlyLysAlaMetSerSerThrGluG 595
 Db 759 ATAATACGTGGAAACAATAGTCACATAGCGTGGTGGCAAGGCAATGTCTCAACCGAGG 700
 QY 595 luLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIlePheGlyG 615
 Db 699 AGAAGTCTACGAAATATGACATCTCTTGACGGTGGACTACGTTTGGTGATCTTTGGCG 640
 QY 615 lyValIleGlyTyrSerClyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaG 635
 Db 639 GTGTGATCGGTATTTCGCGCATGATATCAACAAGTTGCTGTGGATGGTCCGAATGGCG 580
 QY 635 luGlyGluHisProLysAspIleLysGluSerAspTyrPheThrAspArgGlyGluPheA 655
 Db 579 AGGGAGAGCATCCCAAGGACATTAAAGAAAGCGATTACTTTACCGACCGCGGTGAATTCA 520
 QY 655 rgValAspAlaGluGlyAlaProAlaIleuLeuAsnCysLeuMetTyrLysLeuSerTyrT 675
 Db 519 GGGTAGATGCGGAAGGTGCTCCGCGCTGCTCAACTGCTTATGTACAAATTAAGTACT 460
 QY 675 yrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgA 695
 Db 459 ACAGATTGCGGGAATTGAAGTTGACTACAGAGGCCCATCTCGGATATGATCGCACAGTA 400
 QY 695 snAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrG 715
 Db 399 ACGCGGTCAATTGGGAAATGAAGACTTTGATCTGACCTACCTGGAGGAGGCTACACACAG 340
 QY 715 luHisTrpLeuValArgIleTyrArgValLysLysProHisGluPheAsnArgProSerL 735
 Db 339 AACACTGGCTTGTTCGCATCTATAGTGTGAAGAGCGCGCATGAGTTCAATAGACCATCAC 280
 QY 735 euLysThrLysGluArgThrIleProAlaAsnPheIleSerArgLysAsnSerLysA 755
 Db 279 TGAAGACCAAGAGAGAACGATTCTCCAGCAAACTTCATCTCGAAGAAAGAACTCGAAGC 220
 QY 755 rgArgLysGlyTyrIleArgAsnArgProValValLysGlyLysArgThrLeuLys 774
 Db 219 GTGCAAGGGTACTATACGAAACCGACCGGTGTGTGTGAAGGAAACGAACTTGAA 161

RESULT 7

US-09-270-767-11648
 ; Sequence 11648, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 11648
 ; LENGTH: 1386
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-11648

Alignment Scores:

Pred. No.: 3,89e-107 Length: 1386
 Score: 989.00 Matches: 182
 Percent Similarity: 78.83% Conservative: 60
 Best Local Similarity: 59.28% Mismatches: 61
 Query Match: 24.44% Indels: 4
 DB: 4 Gaps: 2

US-10-028-384-8 (1-774) x US-09-270-767-11648 (1-1386)

QY 18 LeuIleThrPheAlaIleLeuLeuIleAlaIleTrpLeuAlaGlyPheSerSerArgLeuPhe 37
 Db 468 CTGCTCAAGCTGGCCATTCTCATCTCGGACGCGTATTATCATTTGCCACACGCTGTTC 527
 QY 38 AlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAla 57
 Db 528 TCTGTGTGGGATTCGAAGCGTAATCCATGAGTTCCATCCGCTACCTCAACATACCGCAC 587
 QY 58 ThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAla 77
 Db 588 ACGCGGTTCTTGGCGGAGCAGGCGCTTTTACAAGTTCCACAACCTGGTTCCATGACCGCGC 647
 QY 78 TrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSer 97
 Db 648 TGGTATCCCTTGGCGCGCATCATCGCGCGCACCATCTATCCCGGCTGATGTCACTCG 707
 QY 98 GlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCys 117
 Db 708 CGCGCCCTGTACCGCTGATGTGCTGCTCAATGTGACCATCGACATACGGAACGTGTGC 767
 QY 118 ValPheIleAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLys 137
 Db 768 GTCTTCTCGCGCGCTTCTTCTCTCGCTGACCAACCGCTGGTACCTACCCCTCAACAG 827
 QY 138 GluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGly 157
 Db 828 GAGATACACAGCACTGGAGCTGGACTGGTGGCGCGCGCTTTGATATCCATCGTTCCCGG 887
 QY 158 TyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeu 177
 Db 888 TATATCTCTCGATCCGTTGGCGGATCGTACGACAAATGAAGGATCGCCATTTTTCGATG 947
 QY 178 GlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAla 197
 Db 948 CTCCTTCACTTACTATTGTGATCAAGCGGTAAAGACGGCACGATCTTTTGGTCGCT 1007
 QY 198 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIle 217
 Db 1008 ATGTCGCGATTTGGCCCTACTTCTATATGTTCTCTCTGGGGTGGCTATGCTTCTCTGAT 1067
 QY 218 AsnLeuIleProLeuHisValPheValLeuLeuMetGlyArgTyrSerProArgLeu 237
 Db 1068 AACCTAATCCGCTGCAGCTGGCGCTGATGATCAACCGAGCTTCTTCGCACAGGATC 1127
 QY 238 LeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProphe 257
 Db 1128 TACATAGCATACAGCAGCTATACTCGCTCGGCAACCAITCTCTCGATGCGATCTCGT 1187
 QY 258 ValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeu 277
 Db 1188 GTGGATTCCAAACCCATCCAGACTCGAACACATGCTGCGACTGGGAACCTTTGGCGCTG 1247
 QY 278 LeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArg 297
 Db 1248 TGCAGATTACCGCTTTCGTCGACTATCTCGCGTCGCGCATTCGCCAAGGATCACATTCGAT 1307
 QY 298 LysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaVal 315
 Db 1308 CTGCTCTTCAAGAGCTGTGTTTCCAGTGTTTG-----ACTGTGTGTGTTCTCTCGTGGT 1361
 QY 316 ValValLeuThrMetLeuGly 322
 Db 1362 ACCCTGCTCAGCTTACCGGG 1382

RESULT 8

US-09-270-767-1318/c
 ; Sequence 1318, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1318
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1318

Alignment Scores:
Pred. No.: 1,21e-101 Length: 560
Score: 937.00 Matches: 183
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.46% Mismatches: 0
Query Match: 23.16% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-1318 (1-560)

```
QY 347 IleAlaSerValSerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHis 366
DB 558 ATTGCATCCGTCGAGCATCAGCCACCACTTGGTTCTCGTTCTTCTTTGATCTGCAC 499
QY 367 IleLeuValCysAlaPheProValGlyValTyrCysIleLysGlnIleAsnAspGlu 386
DB 498 ATCATGTGTGGCCCTTCCAGTGGAGTGTGGTACTGCTCAAGCATCAAGCAGAG 439
QY 387 ArgValPheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 406
DB 438 CGCGTTTTCGTGTGTGTACGCCCATCAGTGGCGTTTACTTGGCTGTGTGTGGTGGT 379
QY 407 LeuMetLeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeu 426
DB 378 TTGATGTGTACCCCTCAGCGGTGTGTGTGTCATGCTGGCCGAGTGGCTTTTCGGACTG 319
QY 427 LeuAspValPheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAla 446
DB 318 TTGATGTGTCTTCGCAAGAGATTCTCTAAGCGAATGGGCACAGCCATAAGCGCAGCC 259
QY 447 ThrGluValAspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspIleAlaGly 466
DB 258 ACCGAAGTGGATGAAGTGAAGTTCATTAAGCAAGAGACGCTGTACGACAGGCTGGC 199
QY 467 LysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeu 486
DB 198 AAGCTGAAGCATCGTACTAAGCATGTATGCTTAATGCTGTTGATGATGCTCCACTGCACG 139
QY 487 LysSerIleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThr 506
DB 138 AAGAGTATTGTTATTTTGGCCGTTCTAATGCTGTTGATGATGCTGCTCCACTGCACG 79
QY 507 TrpValThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGln 526
DB 78 TGGGTGACCAAGCAATGCTACTCCAGTCCCTCCATTGCTTGGCTTTCCACACAGCTCAA 19
QY 527 AspGlySerArg 530
DB 18 GATGATCCCGC 7
```

RESULT 9

US-09-270-767-16600/c
; Sequence 16600, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 16600

; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16600

Alignment Scores:
Pred. No.: 1,21e-101 Length: 560
Score: 937.00 Matches: 183
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.46% Mismatches: 0
Query Match: 23.16% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-16600 (1-560)

```
QY 347 IleAlaSerValSerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHis 366
DB 558 ATTGCATCCGTCGAGCATCAGCCACCACTTGGTTCTCGTTCTTCTTTGATCTGCAC 499
QY 367 IleLeuValCysAlaPheProValGlyValTyrCysIleLysGlnIleAsnAspGlu 386
DB 498 ATCATGTGTGGCCCTTCCAGTGGAGTGTGGTACTGCTCAAGCATCAAGCAGAG 439
QY 387 ArgValPheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 406
DB 438 CGCGTTTTCGTGTGTGTACGCCCATCAGTGGCGTTTACTTGGCTGTGTGTGGTGGT 379
QY 407 LeuMetLeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeu 426
DB 378 TTGATGTGTACCCCTCAGCGGTGTGTGTGTCATGCTGGCCGAGTGGCTTTTCGGACTG 319
QY 427 LeuAspValPheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAla 446
DB 318 TTGATGTGTCTTCGCAAGAGATTCTCTAAGCGAATGGGCACAGCCATAAGCGCAGCC 259
QY 447 ThrGluValAspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspIleAlaGly 466
DB 258 ACCGAAGTGGATGAAGTGAAGTTCATTAAGCAAGAGACGCTGTACGACAGGCTGGC 199
QY 467 LysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeu 486
DB 198 AAGCTGAAGCATCGTACTAAGCATGTATGCTTAATGCTGTTGATGATGCTCCACTGCACG 139
QY 487 LysSerIleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThr 506
DB 138 AAGAGTATTGTTATTTTGGCCGTTCTAATGCTGTTGATGATGCTGCTCCACTGCACG 79
QY 507 TrpValThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGln 526
DB 78 TGGGTGACCAAGCAATGCTACTCCAGTCCCTCCATTGCTTGGCTTTCCACACAGCTCAA 19
QY 527 AspGlySerArg 530
DB 18 GATGATCCCGC 7
```

RESULT 10

US-09-248-796A-3089
; Sequence 3089, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3089
; LENGTH: 867

TYPE: DNA
ORGANISM: Candida albicans
FEATURE:
NAME/KEY: unsure
LOCATION: (25)
OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown

US-09-248-796A-3089

Alignment Scores:

Pred. No.: 1,378-88 Length: 867
Score: 830.00 Matches: 165
Percent Similarity: 71.94% Conservative: 35
Best Local Similarity: 59.35% Mismatches: 62
Query Match: 20.51% Indels: 16
DB: 4 Gaps: 5

US-10-028-384-8 (1-774) x US-09-248-796A-3089 (1-867)

QY 489 lleValleLeuAlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpVal 508
DB 49 TTGGTTTACTGACATTTTACATTTTACCTTTTCTTACTTTTGTATTTTACATTTGTTGGTA 108
QY 509 ThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly 528
DB 109 ACATCGAATGCTTATTCATCACCATCAGTGTGTTTACCATCCCAAC-----CAGATGCG 165
QY 529 SerArgAsnIleLeuAspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAla 548
DB 166 TCACACATATCATTTGATGATATAGAGAGCCCTATTACTGGTTAAGATGAATACACCA 225
QY 549 AspAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 568
DB 226 GAAGATGCCAAGTTATGCTGCTGGATTTATGTTTATCAATCGGGGTATGGCTGAT 285
QY 569 ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 588
DB 286 AGAACACACACTGTGTATACAAATACATGGAATACACATATTCCTGCTGTGTGAAG 345
QY 589 AlaMetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyr 608
DB 346 GCATGCTTCCCTCGAAGATGCTCGTATGAATTTTGACACACACGATGTTGATTAT 405
QY 609 ValLeuValIlePheGlyValIleGlyTyrSerGlyAspIleAsnLeuPheLeu 628
DB 406 GTGTTAGTATATTGGAGGCTTATTCGGTTATTCGTGATGATATTAACAAATTCCTTA 465
QY 629 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
DB 466 TGGATGGTAAGATTTGCTGAAGTATCTGGCTGATGAATCAAGATCAAGACACTACTTT 525
QY 649 ThrAspArgGlyGluPheArgValAspAlaGluAlaProAlaLeuLeuAsnCysLeu 668
DB 526 ACTGACCGAGAGAGATATAAGTGGATAAAGATGTCATCTGCGCAATGAAGATTTCTTTG 585
QY 669 MetTyrLysLeuSerTyrTrpArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
DB 586 ATGATATAAGTATCGTATCATAGATTCAGTGAATG-----TTTGAGGTAGAGAT 636
QY 689 GlyTyrAspArgThrArgAsnAlaValIle---GlyAsnLysAspPheAspLeuThrTyr 707
DB 637 GGTGTGTAGAGTTAGAGTAAACCAACAAATCCAGCCCAATGAGTACCGAAATTTGAATGTT 696
QY 708 LeuGluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLysPro 727
DB 697 GTTGAGAGAGCTTCACATCAGAAAATGGATGTGAGATTTTCAAGATTAA-----750
QY 728 HisGluPheAsnArgProSerLeuLysThrLysGluArgThrIleProAlaAsnPhe 747
DB 751 -----GATTTGGATAATGTTGGTAGAGATTTTACATCAAGCTACTGCT 792
QY 748 IleSerArgLysAsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765
DB 793 TTGAGAGATCATCATCGGCACCTCCAAA-----AGAACAGATCCATA 837

RESULT 11

US-09-513-999C-1965
Sequence 1965, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.Y.
APPLICANT: Giordano, J.Y.

TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 1965
LENGTH: 507
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 122..505
US-09-513-999C-1965

Alignment Scores:

Pred. No.: 2,538-81 Length: 507
Score: 765.50 Matches: 141
Percent Similarity: 91.72% Conservative: 14
Best Local Similarity: 83.43% Mismatches: 13
Query Match: 18.92% Indels: 1
DB: 4 Gaps: 1

US-10-028-384-8 (1-774) x US-09-513-999C-1965 (1-507)

QY 513 TyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIle 532
DB 2 TACTCTAGTCCAAGTGTAGTCTCTGGCTCATACAAAT---CATGATGGCACCAGGAATATC 58
QY 533 LeuAspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAlaAspAlaArg 552
DB 59 TTGATGATTTTAGAGAAGCTTACTTTTGGCTTAAGCAAAATACAGATGAACATCACGA 118
QY 553 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 572
DB 119 GTAATGCTTGGTGGGATTTATGCTATCAGATAGCTGGATGGCTAATAGAACTACGTTG 178
QY 573 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 592
DB 179 GTGGATAATAACACCTGGGAATACACGCCATAGCACTGTGGGAAAAGCTATGCTCTCT 238
QY 593 ThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIle 612
DB 239 AATGAACACAGCAGCCTATAAATCATGAGGACTCTAGATGTAGATTATGTTTGGTTATT 298
QY 613 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 632
DB 299 TTTGAGAGGGTATTATGCTTATCTGGTGATGATATCAACAAATTTCTCTGGATGTTAGG 358
QY 633 IleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPheThrAspArgGly 652
DB 359 ATAGTCAAGAGAGAACATCCCAAGACATTCGGAAGTGTACTATTTTACCCACAGGGA 418
QY 653 GluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMetTyrLysLeu 672
DB 419 GAATTCGCTGTAGACAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGTATATAAATG 478
QY 673 SerTyrTyrArgPheGlyGluLeuLys 681
DB 479 TCATACTACAGATTTGGAGAAATGCAG 505

RESULT 12

US-09-270-767-12856
 ; Sequence 12856, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:

APPLICANT: Homburger et al.
 TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 FILE REFERENCE: File Reference: 7326-094
 CURRENT APPLICATION NUMBER: US/09270767
 CURRENT FILING DATE: 1999-03-17
 NUMBER OF SEQ ID NOS: 62517
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 12856
 LENGTH: 914
 TYPE: DNA
 ORGANISM: Drosophila melanogaster
 US-09-270-767-12856

Alignment Scores:
 Pred. No.: 7,39e-79 Length: 914
 Score: 748.50 Matches: 138
 Percent Similarity: 79.59% Conservativeness: 18
 Best Local Similarity: 70.41% Mismatches: 33
 Query Match: 18.50% Indels: 7
 DB: 4 Gaps: 2

US-10-028-384-8 (1-774) x US-09-270-767-12856 (1-914)

```

QY 534 AspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAlaAspAlaArgVal 553
DB 5 GATGACTTCGCGAGCCCTACTGCTGCAGATCAACACTCCGAGAGCGCTCGGATA 64
QY 554 MetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThreVal 573
DB 65 ATGTCCTGGTGGAGCTACGGCTACCGATACGCGCATGCGCAATCGGACCATATTAGTG 124
QY 574 AspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerThr 593
DB 125 GATAACAATACTTGGACACACACATATATCGCGCTGGCCAGCGATGGCTCTTCG 184
QY 594 GluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIlePhe 613
DB 185 GAGGAGAAGCGCTACGAGATATGAGGGAACCTGGATGGACTACGTTCTCGTATTC 244
QY 614 GlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgile 633
DB 245 GGAGGCTCCTGCTACTTCATCGGACGATATCAACAGTTCTCTGGATGGCGCAT 304
QY 634 Ala-----GluGlyGluHisProLysAspIleLysGluSerAspTyrPheThr 649
DB 305 GCGCGGACGACGATCGTGTGCGCAC-----ATCGGGAAGAGGACTACTATGCG 355
QY 650 AspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMet 669
DB 356 GCCAACGGAGTTCGGAGTGGACAAGAGGGCTCCACCACTCTCAATTGTTGATG 415
QY 670 TyrLysLeuSerTyrTrpArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGly 689
DB 416 TACAGATGTGCTACTATCGCTTTGGGCAAAATGACACGGAAGGTGGCAAGCCAGGCG 475
QY 690 TyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeuGlu 709
DB 476 TACGATCGATGTGTCGCGCGAGATCGGCAACAGGACTTGAATCGATGTCCTGGAG 535
QY 710 GluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLys 725
DB 536 GAGCGGTACACCGAGACGCTGCTGCTGGTGGCGATCTCAAGGTTAAG 593

```

RESULT 13

US-09-248-796A-3090
 ; Sequence 3090, Application US/09248796A
 ; Patent No. 6747137
 ; GENERAL INFORMATION:

APPLICANT: Keith Weinstock et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
 FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.132
 CURRENT APPLICATION NUMBER: US/09248796A
 CURRENT FILING DATE: 1999-02-12
 PRIOR APPLICATION NUMBER: US 60/074,725
 PRIOR FILING DATE: 1998-02-13
 PRIOR APPLICATION NUMBER: US 60/096,409
 PRIOR FILING DATE: 1998-08-13
 NUMBER OF SEQ ID NOS: 28208
 SEQ ID NO 3090
 LENGTH: 594
 TYPE: DNA
 ORGANISM: Candida albicans
 US-09-248-796A-3090

Alignment Scores:
 Pred. No.: 3,58e-56 Length: 594
 Score: 555.00 Matches: 110
 Percent Similarity: 72.83% Conservativeness: 24
 Best Local Similarity: 59.78% Mismatches: 38
 Query Match: 13.72% Indels: 12
 DB: 4 Gaps: 5

US-10-028-384-8 (1-774) x US-09-248-796A-3090 (1-594)

```

QY 11 LysValAlaGlyTyrSer---SerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeu 29
DB 52 AAAGTATTGGCGATAGATGTTGAACCTGATTAGAGTTTATTAAAGTTATTATTATA 111
QY 30 -----AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIle 45
DB 112 TCGATAGACAGTGGAGCTATTCTCTCGTTTATTTCGGTGATTCGATTGGAAGTATT 171
QY 46 IleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGly 65
DB 172 ATTCATGAATTCGATCCTTGGTTCATTCGAGCAACCAATATTAGTCACTCATTC 231
QY 66 TrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleVal 85
DB 232 TTTTATGAATTTTGAATTGGTTTGAATGATAGTAGACTGGTACCCATGGGAGAGTCACT 291
QY 86 GlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHis 105
DB 292 GGTGTACTTTATATCCGGTTTAATGGTGACTTCAGTGCCCAATT---TGG-----CAT 342
QY 106 ValLeuAsn-----IleProValHisIleArgAspIleCysValPheLeuAla 121
DB 343 ATTTTACGTGATTGGTTGGCTTACCGTTGATATATAATAATTGTGTTTATTAGCA 402
QY 122 ProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrpSer 141
DB 403 CCAGTTTTCGGGATTAACCTGCAATTTGACTTATTTTGTACTAAAGAAATGAACGAT 462
QY 142 AlaGlyAlaGlyLeuPheAlaAspPheIleAlaIleValProGlyTyrIleSerArg 161
DB 463 TCTAGTGCACATATTGGCAGCTATTATTATGGGATTCGCCAGGTATATTTCAAGA 522
QY 162 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 181
DB 523 TCAGTGGCTGGTTCTTATGATATGAAGCAATTCGCCATTACTTTATTATGCGCAACATTT 582
QY 182 PheLeuTrpVal 185
DB 583 TATTCTGGATT 594

```

RESULT 14

US-09-328-111-332
 ; Sequence 332, Application US/09328111
 ; Patent No. 6262333
 ; GENERAL INFORMATION:
 ; APPLICANT: Endege, Wilson O.

APPLICANT: Steinmann, Kathleen E.
APPLICANT: Aetle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Derti, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Margia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
TITLE OF INVENTION: PRODUCTS
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328.111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 332
LENGTH: 616
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) - (616)
OTHER INFORMATION: n = A,T,C or G

US-09-328-111-332

Alignment Scores:

Pred. No.:	4,38e-56	Length:	616
Score:	554.50	Matches:	112
Percent Similarity:	82.8%	Conservative:	14
Best Local Similarity:	73.6%	Mismatches:	26
Query Match:	13.70%	Indels:	1
DB:	3	Gaps:	0

US-10-028-384-8 (1-774) x US-09-328-111-332 (1-616)

QY	155	ValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIle	174
DB	5	GTACACAGCTACATATCTCGTCAGTAGCTGGATCCCTTGTATATGAGGCGTGTATT	64
QY	175	PheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPhe	194
DB	65	TTTGCACTTCAGTTCCACATATTTATGGGTAAATCTGTAAACTGGGTGAGTTTTT	124
QY	195	TrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrVal	214
DB	125	TGGACAATGTGCTGCTTATCTTATCTTATATGCTCTGCTTGGGGTGTATGTA	184
QY	215	PheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSer	234
DB	185	TTTATCATCAATCTTATTTCCATGCTATTTTGTGTGTGTGTGTGTGTGTGTGTGT	244
QY	235	ProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGln	254
DB	245	AAAAGAGCTACATAGATATAGCATCTTCTACATTTGGGTTTATATTAATCAATGCAG	304
QY	255	IleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaLeuGlyVal	274
DB	305	ATACCTTTTGTGGGATTCAGCAATCAAGAAAGTGAACACATGGCAGCTTCAGGTGC	364
QY	275	PheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsn	294
DB	365	TTTGCAATGCTGCACACCTTATCTTCTTGTGAGTATCTGAGAACCGA-TTACCAACCAA	423
QY	295	GluPheArgLysLeuPheIleValGlyLeuLeu	306
DB	424	GAGTTCCAGACCTTTCTCTTTTGGGGGATACTACTT	459

RESULT 15

US-09-270-767-27262
Sequence 27262, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 27262
LENGTH: 503
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-27262

Alignment Scores:

Pred. No.:	1.37e-50	Length:	503
Score:	507.00	Matches:	99
Percent Similarity:	76.19%	Conservative:	29
Best Local Similarity:	58.93%	Mismatches:	36
Query Match:	12.53%	Indels:	4
DB:	4	Gaps:	2

US-10-028-384-8 (1-774) x US-09-270-767-27262 (1-503)

QY	157	GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla	176
DB	2	GGGTATATCTCGATCCGTGGGGGATCTGTACGACATGAGGATCCCATTTCTGTC	61
QY	177	LeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer	196
DB	62	ATGCTCTTCACTACTATTGTGATCAAGCGGTAAAGACGGCAGCATCTTTGTCG	121
QY	197	AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle	216
DB	122	GCTATGTGGCAATGGCTACTTCTATATGCTCTCTGTTGGGTGGCTATGCTTCTG	181
QY	217	IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg	236
DB	182	ATTAACCTTAATCCCGTGCACGTGCTGGCGCTGATGATCACCGACGTTTCTGCACAG	241
QY	237	LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro	256
DB	242	ATCTACATAGCATACAGCAGCTATCTGCTGGGACCAATCTGTCGATGCAGATCTCG	301
QY	257	PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal	276
DB	302	TTTGTGGGATTCACACCCATCCAGAGCTCCGACACATGCTGCGCACTGGGAACCTTTGGC	361
QY	277	LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe	296
DB	362	CTGTGCCAGATTCACGTTTCTGTCGACTATCTGCTGGCGCATTCACAGGATCACTTC	421
QY	297	ArgLysLeuPhe-----IleValGlyGlyLeuValGlyValGlyValPheValAla	314
DB	422	GATCTGCTCTTCAAGACGTTGGTTCCAGGTTTGG-----ACTGTGGTGTTCGTCG	475
QY	315	ValValValLeuThrMetLeuGly	322
DB	476	GGTACCTGCTCAGCTTACCGGG	499

Search completed: December 15, 2004, 11:59:40
Job time : 190.55 secs

Plant
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 15, 2004, 06:02:06 ; Search time 974.477 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 4046

Sequence: 1 MNRTPKMLNKKVAGYSLIT.....RRKGYIRNPVVVKGKRLK 774

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131523_8140/app_query.fasta_1.3740
-DB=Published Applications NA -QWTF=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPL=0 -COOPEXT=0 -UNITS=bits -START=1 -END=4 -1 -MATRIX=blomch2
-TRANS=human40 cdi -LIST=45 -DOALIGN=200 -THR SCORE=ext -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10028384@cgn_1_1_2577@runat_14122004_131523_8140
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	4046	100.0	2417	15	US-10-028-384-7
2	3949	72.9	2484	15	US-10-028-384-1
3	3948	72.9	2710	15	US-10-028-384-3
4	2198.5	54.3	3093	18	US-10-417-375-92
5	2198.5	54.3	3094	15	US-10-028-384-9
6	2194.5	54.0	3046	18	US-10-417-375-95
7	2185.5	54.0	5404	18	US-10-417-375-99
8	2185.5	54.0	5827	18	US-10-417-375-97
9	2177.5	53.8	2472	15	US-10-171-581-112
10	2177.5	53.8	2472	15	US-10-028-384-11
11	2177.5	53.8	2472	15	US-10-172-118-742
12	2177.5	53.8	2472	16	US-10-342-887-742
13	2176	53.7	2256	15	US-10-032-585-6323
14	2172.5	53.7	2232	15	US-10-128-714-7139
15	2113.5	52.2	2466	16	US-10-320-797-2305
16	2068	51.1	2681	16	US-10-425-114-14408
17	2060.5	50.9	2779	17	US-10-437-963-99904
18	2059.5	50.9	2839	18	US-10-425-115-150745
19	2036.5	50.3	2157	9	US-09-801-368-387
20	2036.5	50.3	2157	18	US-10-793-639-318
21	2027.5	50.1	2733	15	US-10-028-384-5
22	2014	49.8	2603	15	US-10-128-714-6139
23	2014	49.8	3969	15	US-10-128-714-139
24	2014	49.8	4603	15	US-10-128-714-5139
25	1898.5	46.7	1848	15	US-10-128-714-2139
26	1861.5	46.0	3141	18	US-10-425-115-130787
27	1853	45.8	3197	16	US-10-424-599-111541
28	1851.5	45.8	2882	16	US-10-320-797-1305
29	1851.5	45.9	4738	16	US-10-320-797-305
30	1840.5	45.5	1969	15	US-10-128-714-1139
31	1608.5	39.8	2244	17	US-10-437-963-48342
32	1512	37.4	1543	10	US-09-974-879-133
33	1512	37.4	1543	10	US-09-305-736-133
34	1512	37.4	1543	10	US-09-818-683-133
35	1512	37.4	1543	11	US-09-818-683-133
36	1512	37.4	1543	16	US-10-621-481-133
37	1509.5	37.3	1209	15	US-10-106-698-330
38	1509.5	37.3	1209	16	US-10-264-237-412
39	1282	31.7	1828	10	US-09-945-527-62
40	1231	30.4	1728	16	US-10-424-599-122476
41	1147.5	28.4	2660	16	US-10-264-049-630
42	1006.5	24.9	1114	16	US-10-296-115-629
43	973	24.0	1094	17	US-10-437-963-99902
44	750.5	18.5	887	16	US-10-424-599-77697
45	727	18.0	500	9	US-09-998-598-1643

ALIGNMENTS

RESULT 1
US-10-028-384-7
; Sequence 7, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1.
; SEQ ID NO 7
; LENGTH: 2417
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AF132552
; DATABASE ENTRY DATE: 1999-04-27

RELEVANT RESIDUES: (1)...(2417)
US-10-028-384-7

Alignment Scores:

Pred. No.: 0 Length: 2417
Score: 4046.00 Matches: 774
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-8 (1-774) x US-10-028-384-7 (1-2417)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20
DB 71 ATGAATCGGACGCGGAGATGCTGAACACAGAGGTGGCTACAGAGCCTAATCACC 130
QY 21 PheAlaIleLeuLeuLeuAlaTptLeuAlaGlyPheSerSerArgLeuPheAlaValle 40
DB 131 TTGGCCATCTCGTAAATCGCTGGCTGGCGGATTTCTCTCGCTCTCTCGCTCATC 190
QY 41 ArgPheGluSerIleIleHisGluPheAspProTptPheAsnTyrArgAlaThrAlaTyr 60
DB 191 CGTTTCGAGTCGATTATCCATGAGTTTGATCCGTGGTTCAACTACCGGGCCACCGCCTAC 250
QY 61 MetValGlnAsnGlyTptTyrAsnPheLeuAsnTptPheAspGluArgAlaTptTyrPro 80
DB 251 ATGGTCGAGATGGTTGGTACACTTCCTCACTGGTTCGACGAGCGCATGTATCGG 310
QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
DB 311 CTCGGCAGGATTCTGGCGGCTACCGTCTATCCCGCCTGATGATTACGTCGGCGGAATC 370
QY 101 HisTptLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
DB 371 CATTGGCTGTGCACGCTACTCAACATACCGGTCATATTCGTGACATCTGGCTGTCTG 430
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuLeuThrLysGluLeuTpt 140
DB 431 GCGCGCATCTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGTGG 490
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
DB 491 TCCGCGGGCCGGCTCTTCGCGCCAGCTTCATCGCATCGTGGCTGGCTATCAGT 550
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
DB 551 AGSTCGGTGGCTGGATCGTACGATACAGAGGCGATTGCCATATTCCGCTCGCATTCACC 610
QY 181 TyrPheLeuTptValArgSerValLysThrGlySerValPheTptPheSerAlaAlaAla 200
DB 611 TACTTCTGGTGGTGGCTCAGTGAAGACTGGATCGGTCTGGTGGCGCGAGCGCT 670
QY 201 LeuSerTyrPheTyrMetValSerAlaTptGlyTyrValPheIleIleAsnLeuIle 220
DB 671 TTGTCTCTACTTACATGATGTCGCGCTGGGTGGCTACGTTTCATCATCAACCTGATA 730
QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
DB 731 CCCCTGCACGCTTCTGCTACTGCTCAATATGGGCGAGTACTCGCGCGTCTCTGCACCGC 790
QY 241 TyrSerThrPheTyrIleLeuGlyLeuPhePheSerMetGlnIleProPheValGlyPhe 260
DB 791 TACAGCACCTTCTACATCTGGGACTGCTGTCTCCATGCGAGATCCCTCTCGTGGGATC 850
QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
DB 851 CAACCGATACGACACGAGTGAACACATGCTGGCTGGGAGTGTGTGTCTCTCTATGGCC 910
QY 281 ValAlaThrLeuArgHisIleGlnSerValLeuSerArgAsnGluPheArgLeuPhe 300
DB 911 GTGGCCACCTTCTGGCCATTGACGTCGCTGTCTGGCGAAGAGTTCGCGAAGCTGTCT 970

QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
DB 971 ATCGTCCGCGGATTTGCTGGTGGCGTTGGGCTCTTTGGCGCTGCTGGTCTCACCATG 1030
QY 321 LeuGlyValValAlaProTptPheSerGlyArgPheTyrSerLeuTptPheThrGlyTyrAla 340
DB 1031 CTGGCGCTTTGGGCGCGGTGGAGTGGAGCTTTCTACTCGCTGGGATACTGGCTACGCC 1090
QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrThrPheSer 360
DB 1091 AAGATCCACATTCCTCATTCATTCGTCGAGCATCAGCCACCATCTGGTTCG 1150
QY 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTptTyrCysIle 380
DB 1151 TTCTTCTTTGATCTGCACATCTCTGGTGGCTTCCAGTGGAGTGGTACTGCATC 1210
QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
DB 1211 AAGCAGATCAACGACGAGCGGCTTTCTGGTGTCTGTACGCCATCAGTGGGTTACTTC 1270
QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
DB 1271 GCTGTGTGATGGTGGCTTTGATGTTGACCTCACGCGGTGGTGTGATGTCGCGGA 1330
QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
DB 1331 GTGGCTTTTTCGGGACTGTGGATGTCTCTGCAAGAGGATTCGTCGAAGCATGGGC 1390
QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThr 460
DB 1391 ACAGCATTAAGCGGACGACCGAAGTGGATGAAGCTGAGAGTTCATTCAGAGAGAGAG 1450
QY 461 LeuTyrAspLysAlaGlyLysLeuLysHisIleArgThrLysHisAspAlaGlnGlnAspThr 480
DB 1451 CTGTACGACAAAGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCCGAGGATACT 1510
QY 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuMetMet 500
DB 1511 GCGGTCAGCTCCAACTCCAGAGATTTGTTATTTTGGCGTCTTAATGCTGTGTGATG 1570
QY 501 PheAlaValHisCysThrTptValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
DB 1571 TTGCTGTCTCCTCAGCTGGGTGACCAAGTCTACTCTCCAGTCCCTCCATTCCTTG 1630
QY 521 AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspPheArgGluAlaTyr 540
DB 1631 GCTTTCCAAACAGTCAAGATGGATCCGCAACATTTTAGACATTTAGAGAGGCTTAC 1690
QY 541 TyrTptLeuSerGlnAsnThrAlaAspAlaArgValMetSerTptTptAspTyrGly 560
DB 1691 TACTGGCTTTCCGAGAACATCTGCCGATGATGCTCGGTTATGCTTGGTGGGATACGGA 1750
QY 561 TyrGlnIleAlaGlyMetAlaAsnArgThrLeuValAspAsnSerThrPheAsn 580
DB 1751 TACAGATAGCGGAATGGCAACAGAACACGCTAGTGGATTAATAATACGTGGACAAAT 1810
QY 581 SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle 600
DB 1811 AGTCACATAGCGCTGGTTGGCAAGCAATGCTTCAACGAGGAGAGTCTCCTACGAAAT 1870
QY 601 MetThrSerLeuAspValLeuValLeuValIlePheGlyValIleGlyTyrSer 620
DB 1871 ATGACATCTCTTACGTGGACTACGTTTGTGTGATCTTTGGCGGTGTGATCGGCTATTCT 1930
QY 621 GlyAspAspIleAsnLysPheLeuTptMetValArgIleAlaGluGlyGluHisProLys 640
DB 1931 GCGCATGATATCAACAAGTTCCTGTGGATGCTCCGAATTCGTGAGGAGAGCATCCCAAG 1990
QY 641 AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGly 660
DB 1991 GACATTAGGAAGAGGATTAATTTACCGACCGCGGTGAAATTCAGGCTAGATCCGAGGT 2050
QY 661 AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680

```

Db 2051 GCTCCGGCCCGCTCAACTGCTTATGTACAAATTAAGCTACAGATTCCGGGAATTG 2110
Qy LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
Db 2111 AAGTTGGACTACAGAGTCCATCTGGATATGATCGCACAGTAAACCGCTCATGGGAAT 2170
Qy LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrGluHisTrpLeuValArg 720
Db 2171 AAGGACTTCGATCTGACCTACCTCGAGAGGCGCTCACACAGAACACTGGCTTGTGCG 2230
Qy LysTyrArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
Db 2231 ATCTATAGGGTGAAGAGCGCGCATGAGTCAATAGACCACTCACTGAAGACCAAGGAGAGA 2290
Qy ThrIleProProAlaAsnPheIleSerArgLysAsnSerLysArgLysGlyTyrIle 760
Db 2291 ACGATTCTCCAGCAAACTTCATTTCGAGAAAGAACTCTAAGCGTCGCAAGGCTACATA 2350
Qy ArgAsnArgProValValValLysGlyLysArgThrLeuLys 774
Db 2351 CGAAACCGACCGGTGTTGTTAAGGGAAACGAACCTTGAAA 2392

```

RESULT 2

```

US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)
; OTHER INFORMATION:
US-10-028-384-1

```

```

Alignment Scores:
Pred. No.: 1,74e-293 Length: 2481
Score: 2949.00 Matches: 556
Percent Similarity: 83.88% Conservative: 89
Best Local Similarity: 72.30% Mismatches: 112
Query Match: 72.89% Indels: 12
DB: 15 Gaps: 7

```

US-10-028-384-8 (1-774) x US-10-028-384-1 (1-2481)

```

Qy 10 SerLysValAlaGlyTyrSerSerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeu 29
Db 187 TCGAGCGCGGTGGGTGGCGAGTGGCTTCTCTCTCCATCCATCTCTTCCGCTGGCTT 246
Qy 30 AlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 49
Db 247 GCGCGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAAAGCAATCATCCAGAGTTC 306
Qy 50 AspProTrpPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPhe 69
Db 307 GACCGGTGTTTAACTATAGATCAACATCATCTGCTTCATCTCATCTGCTTCTATGAATTT 366
Qy 70 LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyThrVal 89
Db 367 TTAATTCGTTTGAATGAAGAGCGATGATTCATCCATAGAGAGATAGTAGTGTACTGTT 426

```

```

Qy 90 TyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIle 109
Db 427 TACCCAGGGTGTATGATAAACCCTGGCTTATTCATTTGATTTAAATACATTAACATA 486
Qy 110 ProValHisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer 129
Db 487 ACTGTTTACATATAGAGACGTATGTGTCTTTCCTTGCACCACTTTTAGCGCCTTACACT 546
Qy 130 IleSerThrTyrLeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaIle 149
Db 547 ATATCTACTTCTGCTTACAGAGAACTTGGAAACCAAGAGGAGGAGACTTTTAGCTGCT 606
Qy 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 169
Db 607 TGTTTATTCCTATTGACAGGCTACATATCTCGTCACTAGTGGATCTTCTGATAT 666
Qy 170 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLys 189
Db 667 GAAGGCACTTCTATTTTGCACCTTCAGTTTCACTACTATTTATGGGTAAATCTGTAAA 726
Qy 190 ThrGlySerValPheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAla 209
Db 727 ACTGGGTCAGTTTGGACAATGTCTGCTCTATCTTCTATATGCTCTCTGCT 786
Qy 210 TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIle 229
Db 787 TGGGGTGTATGATTTATTCATCAATCTTATTCACATGCTCATTTATTTGTGTGTTCTG 846
Qy 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeu 249
Db 847 ATGCAGAGATACAGCAAAAGAGCTCACATAGCATATAGCACTTTCTACATTGGGTGTTA 906
Qy 250 LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
Db 907 ATATTATCAATGCAGATACCTTTTGGGATTTCCAGCCAAATCAGAACCAAGTGAACACATG 966
Qy 270 AlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSer 289
Db 967 GCAGCTGCAGTGTCTTTCATTTGCTGCAAGCTTATGCTTTCTGCAGTATCTGAGAGAC 1026
Qy 290 ValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyGlyLeuLeuValGlyVal 309
Db 1027 CGATTAAACAAACAGAGATTCCAGACCTTTTCTTTTGGGTGTATCATCTAGCTGACAGT 1086
Qy 310 GlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGly 329
Db 1087 GCTGTGTCCTTAGTGTCTATTTGACTTATACAGGTATCATTCACCATGGAGTGGC 1146
Qy 330 ArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSer 349
Db 1147 AGGTTTATTCATTTGGGATACCTGGGTATGCAAAAATACACATTTCCAATTATTGCATCA 1206
Qy 350 ValSerGluHisGlnProThrThrTyrPheSerPhePheAspLeuHisIleLeuVal 369
Db 1207 GTGCTGAGCATCACTACGACTTGGGTGCTTCTTCTTCTTGTATCTACATATCTTGTGA 1266
Qy 370 CysAlaPheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPhe 389
Db 1267 TGTACCTTCCAGCAGCGCTTGGTTCCTGCATCAAAAATATCAACGATGAAGAGATATT 1326
Qy 390 ValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 409
Db 1327 GTTGTCTATATGCATCAGTGTCTACTTTGCTGGAGTGTATGGTGGCATGTATGTG 1386
Qy 410 ThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuAspVal 429
Db 1387 ACTTTGACTCCAGTGTGTATGCTGTGCTCAATTTGCCATTTTCAAAATGTTTTTTCAGCAC 1446
Qy 430 PheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluVal 449
Db 1447 TATTTGGGGATGAC---ATGAAAGGGGAAATCCACCTGTGAGGAGCAGCATGATGAG 1503
Qy 450 AspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeu--- 468

```

```

1504 GAT-----GACAAAAGAACCAAGAAATTTGTATGATAGGCGAGGTAAAGTGAGG 1554
Qy LyHisArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSer 488
1555 AAACATGCGCACTGAACAGGAAAAAAGCTGAAGAG--GGATTAGGCGCCCTAATATAAAAGC 1611
Qy IleValIleLeuAlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpVal 508
1612 ATTGTCAACATGTTGATGCTGATGCTATTGATGATGTTTGTGTCCTACCTGGCTC 1671
Qy ThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly 528
1672 ACAAGCAATGCCCTACTCTAGTCCAGTGTAGTCCCTGGCTCATCAAAAT---CATGATGGC 1728
Qy SerArgAsnIleLeuAspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThAla 548
1729 ACCAGGAATATCTTAGATGATTTAGAGAGCTTACTTTGGCTAAGCCAAAATACAGAT 1788
Qy AspAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 568
1789 GAACATGCGCAGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGAAATGCTAAT 1848
Qy ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 588
1849 AGACTACGTTGGTGATGATTAACCTGGATATACACCCATACAGCCATAGCACTGGTGGAAA 1908
Qy AlaMetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyr 608
1909 GCTATGCTCTTCTAATGAACACAGCAGCTATAAAATCATGAGCACTCTAGATGTAGATTAT 1968
Qy ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 628
1969 GTTTGGTTATTTTGGAGGGTTATTTGGCTATTCTGGTGATGATATCAACAAATTTCTC 2028
Qy TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
2029 TGGATGTTAGGATAGCTGAAGAGAACATCCCAAGACATTCGGGAAAGTGACTATATTT 2088
Qy ThrAspArgGlyGluPheArgValAspAlaGluGlyValProAlaLeuLeuAsnCysLeu 668
2089 ACCCCACAGGAGAAATTCGTGTAGACAAAGCAGAGATCCCTACTTTCTGAAATGGCTT 2148
Qy MetTyrIleLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
2149 ATGTATAAAATGTCATCTACTACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCCA 2208
Qy GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
2209 GATTTTGACCGACACAGTAATGCTGATGATGGAATAGGACATTAATTAACAAATTTG 2268
Qy GluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLysPheHis 728
2269 GAAGAAGCCCTTACATCAGAACACTGGCTTGTAGGATATATAAGATAAAAGCACCTGAT 2328
Qy GluPheAsnArgProSerLeuLysThrLysGluArgThr-----IleProProAla 745
2329 -----ACAGGAGAGACATTAGATACAAACCTCGAGTCACCAACATTTCCCAAAACAG 2382
Qy AsnPheIleSerArgLysAsnSerLysArgLysGlyTyrIleArgAsnArgProVal 765
2383 AAGTATTGTCAAAGAGACTACCAAAAGAAAGCGTGGCTACATTAATAAAATAAGCTGGTT 2442
Qy ValValLysGlyLysArgThrLeuLys 774
2443 TTTAAGAGGCGCAAGAAATATCTAAG 2469

```

RESULT 3

US-10-028-384-3

; Sequence 3, Application US/10028384

; Publication No. US20030148285A1

; GENERAL INFORMATION:

; APPLICANT: COMPATIGENE INC.

```

APPLICANT: PERREAULT, Claude
APPLICANT: MCBRIDE, Kevin
TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
FILE REFERENCE: 5600-74
CURRENT APPLICATION NUMBER: US/10/028,384
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn, version 3.1
SEQ ID NO 3
LENGTH: 2710
TYPE: DNA
ORGANISM: Mus musculus
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: AK018758
DATABASE ENTRY DATE: 2001-07-05
RELEVANT RESIDUES: (1)..(2469)
US-10-028-384-3

```

Alignment Scores:

```

Pred. No.: 2,53e-293 Length: 2710
Score: 2948.00 Matches: 551
Percent Similarity: 83.14% Conservative: 90
Best Local Similarity: 71.47% Mismatches: 114
Query Match: 72.86% Indels: 16
DB: 15 Gaps: 5

```

US-10-028-384-8 (1-774) x US-10-028-384-3 (1-2710)

```

Qy 10 SerLysValAlaGlyTyrSerSerLeuIleThrPheAlaIleLeuLeuLeuLeuLeu 29
Db 254 TCGCAGCGCGCGCGGCGGCGAGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 313
Qy 30 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 49
Db 314 GCGCGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAGAGCATCATCCACAGTTC 373
Qy 50 AspProTyrPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPhe 69
Db 374 GACCGCGTGTGTTAACTATAGATCAACACATCATCTTGCATCTCTCATGATTTCTATGATTT 433
Qy 70 LeuAsnTrpPheAspGluArgAlaTyrTyrProLeuGlyArgIleValGlyGlyThrVal 89
Db 434 CTAAATGCTGTTGATGAAGAGCATGTGATCCACTGGAAGATAGTGGGTGGCACCGTT 493
Qy 90 TyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuHisValLeuAsnIle 109
Db 494 TACCCAGCGTTGATGATAACAGCTGGCGCTTATTTCATTGGATTTTAAATACATTGAACATA 553
Qy 110 ProValHisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer 129
Db 554 ACAGTTTCATCAAGAGATGTGTGTATCTCTGCACCACTTTTAGCGGCTTATCATCC 613
Qy 130 IleSerThrTyrLeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAla 149
Db 614 ATATCTAGCTTCTCTGCTAACTAGAGAACTGTGAACCAACAGGACGAGCTTCTAGCTGCT 673
Qy 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 169
Db 674 TGCITCANTGCTATCTACAGGATACATATCTCGGTACGTGGCGGATCTTTGATAAT 733
Qy 170 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLys 189
Db 734 GAAGGCAATGGCATTTTGGCGCTTCACTTACTTACTTATGGGTAAAGTCTGTAAG 793
Qy 190 ThrGlySerValPheTrpSerAlaAlaLeuSerTyrPheTyrMetValSerAla 209
Db 794 ACCGGGTCTGTGTTCTGACAAATGTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGCG 853
Qy 210 TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIle 229
Db 854 TGGGAGGTTATGTGTTTCATCATCACTCATCCCTCTCCATGTGTGTGTGTGTGTGTGTGT 913
Qy 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeu 249

```


Db 914 ATGCAGAGTACAGACAGAGTCTATACATCATATAGCATTTGTACATTTGGGGTTTA 973
Qy 250 LeuPheSerMetGlnProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
Db 974 ATATTATCCATGACAGATACCTTTTGTGGGATTTTCAGCCATCATCAGAACAGGACACATG 1033
Qy 270 AlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSer 289
Db 1034 GCAGCTGCAGGTGCTTTTGCCTGCTCAGAGCTTACGCTTTTGTGAGTATCTGAGAGAC 1093
Qy 290 ValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyGlyLeuValGlyVal 309
Db 1094 CGGTGACAAACAGGAGTCCAGACCTTTTCTTTTGGGTCTCTACCTAGCTGACGAC 1153
Qy 310 GlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGly 329
Db 1154 GCTGTGTTTCCCTAGTGTCATCTATCTGACATACACAGTTATATGACCATGGAGTGGC 1213
Qy 330 ArgPheTrpSerLeuTrpAspThrGlyTrpAlaLysIleHisIleProIleAlaSer 349
Db 1214 AGGTTTATTCATGATGATCTGATGATGATGATGATGATGATGATGATGATGATGAT 1273
Qy 350 ValSerGluHisGlnProThrTrpPheSerPhePhePhePhePhePhePhePhePhe 369
Db 1274 GTGCTGAACATCAGCTTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1333
Qy 370 CysAlaPheProValGlyValTrpCysIleLysGlnIleAsnAspGluArgValPhe 389
Db 1334 TGTACCTTCCAGCAGGCTATGCTTCTGATCAAAATATCAACGATGAAGATATTT 1393
Qy 390 ValValLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetValLeuMetLeu 409
Db 1394 GTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1453
Qy 410 ThrLeuThrProValValCysMetLeuAlaValAlaPheSerGlyLeuLeuAspVal 429
Db 1454 ACTCTGACCCGGTCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1513
Qy 430 PheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluVal 449
Db 1514 TATTTGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1552
Qy 450 AspGluAlaGluAspSerIleGluLysLys-----ThrLeuTrpAspLysAlaGly 466
Db 1553 CAGGACAGCAGTGTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1612
Qy 467 LysLeuLysHisArgThrLysHisAspAlaGlnAspThrGlyValSerSerAsnLeu 486
Db 1613 AAAGTGAAGAGCATGTGACAGACCAAGAACCTGAGAGGCTTTGGGCCCAACATC 1672
Qy 487 LysSerIleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThr 506
Db 1673 AAAAGCATGTGACCATGCTGATGCTGATGCTGATGATGATGATGATGATGATGATGATGAT 1732
Qy 507 TrpValThrSerAsnAlaTrpSerProSerIleValLeuAlaPheHisAsnSerGln 526
Db 1733 TGGGTCAACACACCGCTATCTCAGTCCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789
Qy 527 AspGlySerArgAsnIleLeuAspAspPheArgGluAlaTrpTrpLeuSerGlnAsn 546
Db 1790 GATGTTACAGCAATATATTAGATGATTTTAGAGAGCGTACCTTTTGGCTGAGACAAAC 1849
Qy 547 ThrAlaAspAspAlaArgValMetSerTrpTrpAspTrpGlyTrpGlnIleAlaGlyMet 566
Db 1850 ACGGATGAACACCGCCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1909
Qy 567 AlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaVal 586
Db 1910 GCCAACAGGACCACTCTGTTGGATACACACCTGGAACACACGACCCACATCGCAGTGGTC 1969
Qy 587 GlyLysAlaMetSerSerThrGluLysSerTrpCysIleMetThrSerLeuAspVal 606

Db 1970 GGAAAGCTATGTCTTCCCAATGAACCGCGGCTATATAAATCATGAGGTCCCTTGTATGTC 2029
Qy 607 AspTrpValLeuValIlePheGlyGlyValIleGlyTrpSerGlyAspAspIleAsnLys 626
Db 2030 GATTTATGTGTGTTATTTTCGGAGAGTGAITGGCTATTCGGGGACGATATCAACAAG 2089
Qy 627 PheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAsp 646
Db 2090 TTCTCTGGATGTGAGTAGCTGAAGGGAGCATCCCAAGACATCCGGAGAGTGCAC 2149
Qy 647 TrpPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsn 666
Db 2150 TATTTCCACCCAGCAGGAGAGTTCGAGTAGACAAAGCTGGGTCTCTACTCTGTATAAC 2209
Qy 667 CysLeuMetTrpLysLeuSerTrpTrpArgPheGlyGluLeuLysLeuAspTrpArgGly 686
Db 2210 TGCCTTATGATATAAATGTCATCTACAGATTTGGAGAAATCAGTAGATTTTCGACT 2269
Qy 687 ProSerGlyTrpAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThr 706
Db 2270 CCCCCAGGCTTTGACCCACACGTAATGCTGAGATTGGAAATATAAGACATTAAATTCAAG 2329
Qy 707 TyrLeuGluGluAlaTrpThrTrpLysHisTrpLeuValArgIleTrpArgValLysLys 726
Db 2330 CATTTGGAGGAGCTTTTATCTCGAGCACTGGCTTGTGAGATATATAAGTGAAGCA 2389
Qy 727 ProHisGluPheAsnArgProSerLeuLysThrLysGluArgThr-----IlePro 743
Db 2390 CCTGAC-----AACAGGGAGACACATGAGTTCACAACTCGAGTCCCAACATCGTCCCC 2443
Qy 744 ProAlaAsnPheIleSerArgLysAsnSerLysArgGlyValTrpIleArgAsnArg 763
Db 2444 AACAGAGATTTTGTCAAGAAGACACTATAAAGGAGCGTGGCTACGTTAAAAATAAG 2503
Qy 764 ProValValLysGlyLysArgThrLeuLys 774
Db 2504 CTAGTGTTTAAGAAAGCAAGAGACCTCTAAG 2536

RESULT 4
US-10-417-375-92
; Sequence 92, Application US/10417375
; Publication NO. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 3093
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-417-375-92

Alignment Scores:
Pred. No.: 6,92e-216 Length: 3093
Score: 2198.50 Matches: 416
Percent Similarity: 73.78% Conservative: 113
Best Local Similarity: 58.02% Mismatches: 143
Query Match: 54.34% Indels: 45
DB: 18 Gaps: 8

US-10-028-384-8 (1-774) x US-10-417-375-92 (1-3093)

Qy 17 SerLeuLeuThrPheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeu 36
Db 157 ACATCTTAAGGCTTCTCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 216
Qy 37 PheAlaValIleArgPheGluSerIleHisGluPheAspProTrpPheAsnTrpArg 56

Db 217 TTGCTGCTGCTGAGATTGAAAGTGCATCCATGAGTTGTGATCCGCTACTCTTAATATATCGG 276
Qy 57 AlaThrAlaTyrMetValGlnAsnGlyTyrPheAsnPheLeuAsnTyrPheAspGluArg 76
Db 277 ACTACCCGGTTTCTGGCTGAGGAGGGTTTATAAATTCCTAATCCAACTGCTTTGATGACCGG 336
Qy 77 AlaTyrPheProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
Db 337 GCTTGTACCTTTGGCCGATCATGGAGGACAAATTACCCAGGTTTATGATGACT 396
Qy 97 SerGlyGlyIleHisTyrPheLeuHisValLeuAsnIleProValHisIleAspIle 116
Db 397 TCTGCTCAATCTACCATCTGACTCCATTTCTCCATATCATATTCATGACATTCGAATGTC 456
Qy 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
Db 457 TGTGTTTCTGCCCCACTTTCTCTCTTCCACCCATCGTTACGTACACCTTACC 516
Qy 137 LysGluLeuTyrPheSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
Db 517 AAGAGCTCAAGGATGAGGAGTGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 576
Qy 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
Db 577 GGGTATATTTCTGATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 636
Qy 177 LeuGlnPheThrTyrPheLeuTyrPheValArgSerValLysThrGlySerValPheTyrSer 196
Db 637 ATGCTGCTTACTTACTACATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 696
Qy 197 AlaAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTyrGlyGlyTyrValPheIle 216
Db 697 GCCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 756
Qy 217 IleAsnLeuIleProIleHisValPheValLeuLeuLeuMetGlyArgTyrSerProArg 236
Db 757 ATCAACTGATCTCTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
Qy 237 LeuLeuThrSerTyrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
Db 817 ATCTACGTAGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 876
Qy 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaLeuGlyValPheVal 276
Db 877 TTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 936
Qy 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
Db 937 CTCTGTCAGATCCATGCTTTCGATGATTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 996
Qy 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
Db 997 GAAGTCTTCTCCGAGTGTATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1053
Qy 312 PheValAlaValValLeuLeuMetLeuGlyValValAlaProTyrPheSerGlyArgPhe 331
Db 1054 -----GCTCTCTCATGCTAACA-----GGAAAAATTTCTCCCTGACAGGCGGTTT 1101
Qy 332 TyrSerLeuTyrPheThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
Db 1102 TACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1161
Qy 352 GluHisGlnProThrTyrPheSerPhePheAspLeuHisIleIleValCysAla 371
Db 1162 GAGCACAGCCCAACCTGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1221
Qy 372 PheProValGlyValTyrPheTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
Db 1222 TTTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1281
Qy 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
Db 1282 ATGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1341

Qy 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
Db 1342 GCACCTGTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1401
Qy 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
Db 1402 AAAAATCTGGACATAAAGTCGCCA----- 1425
Qy 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
Db 1426 -----GACAAAGAGCAAG----- 1440
Qy 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
Db 1441 -----AAGCAACAGGATTTCTACCTATTAAAGAAATGAGTGGCGAGTGGG 1488
Qy 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTyrValThr 509
Db 1489 ATGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1548
Qy 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
Db 1549 AGTGAAGCTTATTTCTCTCCCTCCATTTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1605
Qy 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTyrLeuSerGlnAsnThrAlaAsp 549
Db 1606 AGGATCATTTTGTGATGCTTCCGAGAGCGTATTATTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1665
Qy 550 AspAlaArgValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
Db 1666 GATGCAAAAGTCATGTCATGCTGGGATTAAGCTACCAATTTACTCAATGGCAATCGG 1725
Qy 570 ThrThrLeuValAspAsnAsnThrTyrAsnAsnSerHisIleAlaLeuValGlyLysAla 589
Db 1726 ACAATTTTGTGAGCAATAACACATGAATAAACCATATTTCTCGAGTAGGCGAGCA 1785
Qy 590 MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
Db 1786 ATGGCATCCACAGAAAGAAAGCCATGAATCATGAGGAGGCTTGTATGCTAGCTATGTG 1845
Qy 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyr 629
Db 1846 CTGTGCTATTTTGGAGCGCTTACTGGGTATTTCTTGGATGATATCAACAGTTCTTTGG 1905
Qy 630 MetValArgIle---AlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
Db 1906 ATGGTCCGATTTGGAGGAAGCACAGACAGGAGGACACATTAAAGGAGAAATGACTACTAT 1965
Qy 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
Db 1966 ACTCTACTGGGGAATTCGGTGTGATCGTGAGGTTTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 2025
Qy 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
Db 2026 ATGTACAAATGTGTACTACCGCTTTGGCGAGCTCTACACAGAACCAAGCGTCCACCA 2085
Qy 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysPheAspLeuThrTyrLeu 708
Db 2086 GCTTTTGACCTGTTTCCAAATGCTGAGATGGTAAAGACTTTGAGCTTTGATGCTTGTGCTGCTGCTGCT 2145
Qy 709 GluGluAlaTyrThrThrGluHisTyrPheValArgIleTyrArgValLys 725
Db 2146 GAGGAAGGCTATACACAGAACACTGCTGCTAGTCAGGATATACAAAGTAAAG 2196

RESULT 5

US-10-028-384-8
; Sequence 9, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin

[illegible]

```

QY 610 LeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 629
DB 1846 CTTGTCTATTTTGGAGGCTTACTGGGTATTTCTCGATGATGATCAACAAGTTCTTTGG 1905
QY 630 MetValArgIle--AlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
DB 1906 ATGTCGCGGATTGGAGGAGCAGACAGAGAGGAGACATTAAGAGAGTACTACTAT 1965
QY 649 ThrAspArgGlyGluPheArgValAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
DB 1966 ACTCTACTGGGAATTCGCTGTGATCGCTGAGGGTTCTCCGGTCTGCTCAACTGCGCTT 2025
QY 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
DB 2026 ATGTACAAATGTGTACTACCGCTTTGGCAGGTCTACAGAGCCAGCGTCCACCA 2085
QY 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
DB 2086 GCGTTTACCGCTGTTCGAAATGCTGAGATTGGTAAATAAAGACTTTTGAGCTTGATGCTG 2145
QY 709 GluGluAlaTyrThrThrGluHisTyrLeuValArgIleTyrArgValLys 725
DB 2146 GAGGAGCGTATACACAGACAGACTGCTGCTAGTACAGGATATACAGGTAAAG 2196

RESULT 6
US-10-417-375-95
; Sequence 95, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95
; LENGTH: 3046
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-95

Alignment Scores:
Pred. No.: 1.75e-215 Length: 3046
Score: 2194.50 Matches: 417
Percent Similarity: 73.47% Conservative: 112
Best Local Similarity: 57.92% Mismatches: 140
Query Match: 54.24% Indels: 51
DB: 18 Gaps: 9

US-10-028-384-8 (1-774) x US-10-417-375-95 (1-3046)

QY 17 SerLeuIleThrPheAlaIleLeuLeuIleAlaTyrLeuAlaGlyPheSerArgLeu 36
DB 181 ACACCTTTGAGCTTCTCATCTGTCAGATGGTGTGTATATCTCTCTCCACTGCTG 240
QY 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArg 56
DB 241 TTTGCTGCTCCTGAGATTTGAAAGTGTATCCATGATGATTTGATCCCTGCTACTTAAATTCGG 300
QY 57 AlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArg 76
DB 301 ACTACCAAGTCTCTGCTGAGGAGGGGTTTATTAATTCCTACTGTTTGTATGACCGA 360
QY 77 AlaTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
DB 361 GCGTGTACCTTTGGGAGAAATCATTTGAGGAGCAATTAACCAAGGTTTAAATGATCACC 420
QY 97 SerGlyGlyIleHisTyrLeuLeuHisValLeuAsnIleProValHisIleArgAspIle 116
DB 421 TCTGCTGCAATCTACCATGTAATTTTTCACATACCATTCGATTCGGAATGTC 480

```

```

QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
DB 481 TGTGTGTCTCTGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
QY 137 LysGluLeuTyrSerAlaGlyValGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
DB 541 AAGAGAGCTCAAGGATGAGGGCTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
DB 601 GGATATATCTCCCATCTCTGCTGGCTCTCTATGATTAATGAAGGAGTTGCCATCTTTTGC 660
QY 177 LeuGlnPheThrTyrPheLeuTyrValArgSerValLysThrGlySerValPheTyrSer 196
DB 661 ATGCTACTACCTACTACTATGATGATCAAGCAGTAAAGACTGGTTCCTCTCTCTCTCTCT 720
QY 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTyrPheGlyTyrValPheIle 216
DB 721 GCTAAGTGTGGCTTGTCTTATTTCTATGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg 236
DB 781 ATCAACTTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
QY 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
DB 841 ATCTATGTGGCTACTGCTACTGTTTACTGCTGGGCACTATATCTCTCTCTCTCTCTCT 900
QY 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
DB 901 TTTGTGGTGTTCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
DB 961 CTCTGCCAGATCCATGCTCTTGTGATTAATCTCTGCGCAAGTTGAATCCACCAACAATTT 1020
QY 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
DB 1021 GAAGTCTCTTTCCGAGGCTCATCTCTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1077
QY 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProTyrSerGlyArgPhe 331
DB 1078 -----GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1125
QY 332 TyrSerLeuTyrPhePheThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
DB 1126 TACTGCTGCTGGATCCCTCTTATGCTAAGACACACATCCCATCATCTGCTCTGCTCT 1185
QY 352 GluHisGlnProThrTyrPhePhePhePhePhePhePhePhePhePhePhePhePhePhe 371
DB 1186 GAGCATCAGCCCAACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1245
QY 372 PheProValGlyValTyrTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
DB 1246 TTTCCAGTGTGGCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1305
QY 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
DB 1306 ATGTGTGTGTGACGAGCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1365
QY 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
DB 1366 GCACCTGTATGTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1425
QY 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
DB 1426 AAGAATCTCGACATAGTCGTCCTCA-----GACAGAGAGAGCAAG-----1449
QY 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
DB 1450 -----GACAGAGAGAGCAAG-----1464
QY 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489

```

```

Db 1465 -----AAGCACAGATCCACCTACCTATTAGAATGAGTGGCAAGTGGG 1512
Qy 490 VALLLEuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
Db 1513 ATGATACTGGTCATGGCTTCTTCTCATCACTACACCTTCTTCAACCTGGTGACC 1572
Qy 510 SerAsnAlaTyrSerProSerPheValLeuAlaPheHisAsnSerGlnAspGlySer 529
Db 1573 AGTGAGGCTACTTCTTCCGTCCTATGTTACTATCTGCTCCGCTGGTGGG---GATGGCAGT 1629
Qy 530 ArgAsnLeuLeuAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAlaAsp 549
Db 1630 AGGATCATATTGATGACTTCCGAGAAGCATATTATTGGCTTCGTATATCTCCAGAG 1689
Qy 550 AspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnLeuAlaGlyMetAlaAsnArg 569
Db 1690 GATCGGAAGGTCATGCTCGTGGTGGATATGGCTATCAGATTACAGTATGGCAACCGA 1749
Qy 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisLeuAlaLeuValGlyLysAla 589
Db 1750 ACAATTTTAGTGACAATAACACATGGAATAATACCATATTCTCGAGTAGGGCAGGCA 1809
Qy 590 MetSerSerThrGluGluLysSerTyrGluLeuMetThrSerLeuAspValAspTyrVal 609
Db 1810 ATGGCGTCCACAGAGGAAAGGCTATGAGATCATGAGGAGGCTCGATGTCAGCTATGTG 1869
Qy 610 LeuValilePheGlyGlyValileGlyTyrSerGlyAspAspLeuAsnLysPheLeuTrp 629
Db 1870 CTGTCTATTTTGGAGGCTCCTGCGTATCTCTGTGATATATCAACAGTTCTTTGG 1929
Qy 630 MetValArgileAla-----GluGlyGluHisProLysAspLeuLysGluSer 645
Db 1930 ATGGTCGGATTGGAGGAGCACAGATACAGGCAACAT-----ATCAAGGAGAAAT 1980
Qy 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
Db 1981 GACTATATATCTCACTGGGAGTTCCTGTGGAGCCGTGAGGTTCTCCAGTGTGCTC 2040
Qy 666 AsnCysLeuMetTyrLysLeuSerTyrArgPheGlyGluLeuLysLeuAspTyrArg 685
Db 2041 AACTGCCTCATGTACAAGATGTGTACTATCGCTTTGGACAGTTTACACAGAGCCAG 2100
Qy 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValileGlyAsnLysAspPheAspLeu 705
Db 2101 CGTCCCTCCAGCTTTGACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTT 2160
Qy 706 ThrTyrLeuGluAlaTyrThrThrGluHisTrpLeuValArgileTyrArgValLys 725
Db 2161 GATGTCTCGGAGGAGCATATACACAGAACATTTGCTGGTCAGGATATACAGGTAAG 2220

```

RESULT 7

```

US-10-417-375-99
; Sequence 99, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Maiandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 5404
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-99

```

```

Alignment Scores: 3,57e-214 Length: 5404
Pred. No.: 2185.50 Matches: 415
Score:

```

```

Percent Similarity: 73.40% Conservative: 112
Best Local Similarity: 57.80% Mismatches: 140
Query Match: 54.02% Indels: 51
DB: 18 Gaps: 9

```

US-10-028-384-8 (1-774) x US-10-417-375-99 (1-5404)

```

Qy 17 SerLeuileThrPheAlaileLeuLeuileAlaTrpLeuAlaGlyPheSerSerArgLeu 36
Db 181 ACACITTTTGAAGCTTCTCATTCGTCGAATGGCTGTGATTATTCCTTCTCCACTCGTCTG 240
Qy 37 PheAlaValileArgPheGluSerileleHisGluPheAspProTrpPheAsnTyrArg 56
Db 241 TTTGCTGTCTCAGAGATTGAAAGTGTATTCCATGATGTGATCCCTACTCTTAATATATCGG 300
Qy 57 AlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArg 76
Db 301 ACTACCAGGTCTCTCGCTGAGGAGGGTTTATAAATCCATACTGGTTGTATGACCGA 360
Qy 77 AlaTrpTyrProLeuGlyArgileValGlyGlyThrValTyrProGlyLeuMetileThr 96
Db 361 GCCTGGTACCCCTTTGGGAGCAATCATTCGAGAGCAATTTACCCAGGTTTAAATGATCACC 420
Qy 97 SerGlyGlyileHisTrpLeuLeuHisValLeuAsnileProValHisileArgAspIle 116
Db 421 TCTGCTGCAATCTACCATGCTACTCCATTTTTCACATCACCATCGACATTCGGATGTC 480
Qy 117 CysValPheLeuAlaProilePheSerGlyLeuThrSerileSerThrTyrLeuLeuThr 136
Db 481 TGTGTGTTCTCGGCCCTCTCTTCTCTCTTCCCTTCCACCATCGTCACGATCCACCTTACC 540
Qy 137 LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheileAlaileValPro 156
Db 541 AAAGAGCTCAGGATCGAGGGCTGGGCTTCTGCTGCTCCATGATTGCTGTAGTTCCCT 600
Qy 157 GlyTyrileSerArgSerValAlaGlySerTyrAspAsnGluGlyileAlailePheAla 176
Db 601 GGATATATCTCCGATCTGTGGCTGGCTCTATGATATATGAAGGATGCCATCTTTTGC 660
Qy 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
Db 661 ATGCTACTCCTACTACTGATGTGATCAAGCAAGTAAGACTGGTTCATCTGTGTGGCA 720
Qy 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheile 216
Db 721 GCTAAGTGTGCCCCTGCTTATTTCTACATGCTCTCGTCATGGGAGGATTATGTTCTGT 780
Qy 217 IleAsnLeuileProLeuHisValPheValLeuLeuileMetGlyArgTyrSerProArg 236
Db 781 ATCAACTTAATTCCTCTCCAGCTCTCGTGTGATGCTACAGGCCGCTTCTCTCACCGG 840
Qy 237 LeuLeuThrSerTyrSerThrPheTyrileLeuGlyLeuLeuPheSerMetGlnilePro 256
Db 841 ATCTATGTGGCTACTGTACTGTTTACTGCTGGGCACTATCTTTCTATGCAATCTCC 900
Qy 257 PheValGlyPheGlnProileArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
Db 901 TTTGTGGGTTTCCAGCCTGTCTTTCATCAGAGCACATGCGAGCTTTTGGGCTTTTGGT 960
Qy 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
Db 961 CTCTGCCAGATCCATGCTTTGTGATTACTCGCAGCAAGTGAATCCACAACTTT 1020
Qy 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
Db 1021 GAAGTTCTTTTCCGAGCGCTACTCTCTCTGTAGGCTTTTGTCTCTCTCACCGTGGGA-- 1077
Qy 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
Db 1078 -----GCTCTCTCTCATGTGACA-----GGAAAAATATCTCCCTGGAGCGGCTTTC 1125
Qy 332 TyrSerLeuTrpAspThrGlyTyrAlaLysileHisileProileleAlaSerValSer 351

```


QY	217	IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyAcgTyrSerProArg	236
DB	781	ATCAACATAATCTCTCCAGCTCTCGTGTGTGATGCTCAAGGCCGTTCTCTCACCGG	840
QY	237	LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuPheSerMetGlnIlePro	256
DB	841	ATCTATGTGGCTACTGTACTGTATTACTGCTCGGCACATATACTTCTATGCAGATCTCC	900
QY	257	PheValGlyPheGlnProIleArgThrSerGlnHisMetAlaLeuGlyValPheVal	276
DB	901	TTTGTGGTTCAGACCTGCTCTTCATCAGACACATGGCAGCTTTGGGGTCTTTGGT	960
QY	277	LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe	296
DB	961	CTTCGCCAGATCCATGCTTTGTGGATTAACCTCGCAGCAAGTTGAATCCACACAAATTT	1020
QY	297	ArgLysLeuPhe-----IleValGlyLeuLeuValGlyValGlyVal	311
DB	1021	GAAGTTCTTTTCGGAGGCTCATCTCTCTGTAGGCTTTGTCTTCTCACCGTGGGA--	1077
QY	312	PheValAlaValValLeuThrMetLeuGlyValValAlaProTrrpSerGlyArgPhe	331
DB	1078	-----GCTCTCCTCATGCTGACA-----GGAAAAATATCTCCTCGACGGGGCTTTC	1125
QY	332	TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer	351
DB	1126	TACTCGCTGTGGATCCCTCTTATGCTTAAAGAACACATCCCATCATGTGTTCTGTGTCT	1185
QY	352	GluHisGlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla	371
DB	1186	GAGCATCAGCCCAACCTGGTCTCATACTATTTTGACCTGGAGCTCTCTGCTTCATG	1245
QY	372	PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal	391
DB	1246	TTTCCAGTTGGCTCTATTACTGCTTTPAGCAACCTGTCTATGCCCGGATTTTATCATC	1305
QY	392	LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu	411
DB	1306	ATGTATGGTGTACCAGCATGTACTTTACGTGTATGTGTGTCTATGTCTAGTGTGTG	1365
QY	412	ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu	431
DB	1366	GCACCTGTTATGTGCATCTCTCTGGCATGGAGTCTCCAGGTGCTGTCCACATACATG	1425
QY	432	GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu	451
DB	1426	AAGAATCTGCACATAAGTCGTCCA-----GACAAAGAGAGCAAG-----	1449
QY	452	AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg	471
DB	1450	-----AAGCAACAGGATTCACCTACCTCCTATTAAAGTAAGTGAAGTGCACATGGG	1512
QY	490	ValIleLeuAlaValLeuMetLeuLeuMetPheAlaValHisCysThrTrpValThr	509
DB	1513	ATGATACTGGTCTGCTCTCTTCTCATCACCTACACCTTTTCATCCCTGGGTGACC	1572
QY	510	SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer	529
DB	1573	AGTAGGCCCTACTCTCTCCGTCCTATGTACTATCTGCCCGCTGGTGGG---GATGSCAGT	1629
QY	530	ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAsp	549
DB	1630	AGGATCATATTGATGACTTCCGAGAAGCATATTATGGCTTCGTCTATTAATACTCCAGAG	1689
QY	550	AspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg	569
DB	1690	GATGCCAAGGTATGCTCTGGTGGGATTATGGTATGCTATCAGTATCAGCTATGCCAAACCGA	1749

Qy	570	ThrThrLeuValAspAsnAsnThrTyrPasnAsnSerHisIleAlaLeuValGlyIysAla	589
Db	1750	ACAAATTTTAGTGGCAATAACACATGAGTAATAATACCATATTTCTCGAGTAGGCAGGCA	1809
Qy	590	MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal	609
Db	1810	ATGGCGTCCACAGAGGAAAAGACCTATGAGATCATGAGGAGCTCGGATGCAGCTATGTG	1869
Qy	610	LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyr	629
Db	1870	CTGGCTCATTTTGGAGGCGCTCACTGGGTATTCTCTCATGATATCAACAAGTTCTTTGG	1929
Qy	630	MetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluSer	645
Db	1930	ATGTTCGGGATTTGGAGGAGCACAGATACAGGCAAAACAT-----ATCAGGAGCAAT	1980
Qy	646	AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu	665
Db	1981	GACTATTATACTCCAACGTGGGAGTTCCGTGTGGACCGTGAAGTCTTCAGTGTCTGCTC	2040
Qy	666	AsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArg	685
Db	2041	AACTGCCTCATGTACAAGATGTGTTACTATCGCTTTGGACAGGTTTACACAGAAGCCCAAG	2100
Qy	686	GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu	705
Db	2101	CGTCCTCCAGGCTTTGACCGGTGCCGAATGCTGAGATTCGGGAATAAAGACTTTGAGCTT	2160
Qy	706	ThrTyrLeuGluGluAlaTyrThrThrGluHisTyrLeuValArgIleTyrArg	723
Db	2161	GATGCTCTGGAGGAGCATATACACAGAACATTTGCTGTGTGAGTATATACAAG	2214
RESULT 9			
US-10-171-581-112			
; Sequence 112, Application US/10171581			
; Publication No. US2003010426A1			
; GENERAL INFORMATION:			
; APPLICANT: Dai, Hongyue			
; APPLICANT: Linsley, Peter			
; APPLICANT: Mao Mao			
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia			
; FILE REFERENCE: 9301-157-999			
; CURRENT APPLICATION NUMBER: US/10/171,581			
; CURRENT FILING DATE: 2002-06-14			
; PRIOR APPLICATION NUMBER: 60/298,914			
; PRIOR FILING DATE: 2001-06-18			
; NUMBER OF SEQ ID NOS: 366			
; SEQ ID NO 112			
; LENGTH: 2472			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; PUBLICATION INFORMATION:			
; DATABASE ACCESSION NUMBER: L38961			
; DATABASE ENTRY DATE: 2001-06-18			
US-10-171-581-112			
Alignment Scores:			
Pred. No.: 7,22e-214 Length: 2472			
Score: 2177.50 Matches: 415			
Percent Similarity: 72.92% Conservative: 110			
Best Local Similarity: 57.64% Mismatches: 144			
Query Match: 53.82% Indels: 51			
DB: 15 Gaps: 9			
US-10-028-384-8 (1-774) x US-10-171-581-112 (1-2472)			

US-10-028-384-8 (1-774) x US-10-171-581-112 (1-2472)

QY		17	SerLeulleThrPheAlaIleLeuleulealaTrpLeulaGlyPheSerSerArgLeu	36
			:::::	
Db		152	ACACTTTTGAAGCTTCATCTCTCAATGGCTGTCTATTATCCTTCCACTCGTG	211
			:::::	
QY		37	PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrvArg	56
			:::::	
Db		212	TTTGTGTCCTGCAGATTTGAAAGTGTAATCATGAGTTTGATCGTACTTAAATTCGG	271
			:::::	

QY 57 AlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPhelLeuAnThrPheAspGluArg 76
 Db 272 ACTACGAGTTCCTGGCTGAGAGGGGTTTATATAATTCACACTGGTTGATGACCGA 331
 QY 77 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
 Db 332 GCCTGGTACCTTTGGGACGATCATTCGAGGACAAATTTACCCAGGTTTAATGATCAGC 391
 QY 97 SerGlyGlyIleHisTrpLeuHisValLeuAsnIleProValHisIleArgAspIle 116
 Db 392 TCTGCTGCAATTCACCATGTACTCCATTTTCCACATCACCATCGACATTCGGAATGTC 451
 QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
 Db 452 TGTGTGTTCCTGGCCCTCTCTCTCCCTCCATCCATCCATCGTACGCTCCTTACC 511
 QY 137 LysGluLeuTrpSerAlaGlyValaGlyLeuPheAlaIleSerPheIleAlaIleValPro 156
 Db 512 AAAGAGCTCAAGGATGACAGGGCTGGGCTTCTGTCTGCTGCATGATGCTGTAGTTCT 571
 QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
 Db 572 GGATATATCTCCGATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 631
 QY 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
 Db 632 ATGCTACTCACCTACTACATGGATCAAGGAGTAAAGACTGGTTCCATCTGTGGCA 691
 QY 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 216
 Db 692 GCTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 751
 QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg 236
 Db 752 ATCACTTAATTCCTCCACGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 811
 QY 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
 Db 812 ATCTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 871
 QY 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
 Db 872 TTGTGGGTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 931
 QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
 Db 932 CTCTGCCAGATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 991
 QY 297 ArgLysLeuPhe-----IleValGlyGlyLeuValGlyValGlyVal 311
 Db 992 GAAGTTCTTTCCGAGCGTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1048
 QY 312 PheValAlaValValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
 Db 1049 -----GCTCTCTCATGCTGACA-----GGAAAAATATCTCTGCTGCTGCTGCTGCTGCT 1096
 QY 332 TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
 Db 1097 TACTCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1156
 QY 352 GluHisGlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla 371
 Db 1157 GAGCATCAGCCCAACCTGCTCTCATATATTTGACCTGCAGCTGCTGCTGCTGCTGCTGCTGCT 1216
 QY 372 PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
 Db 1217 TTTCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1276
 QY 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
 Db 1277 ATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1336

QY 412 ThrProValValCysMetLeuAlaGlyValaPheSerGlyLeuLeuAspValPheLeu 431
 Db 1337 GCACCTGTTATGACGATTTCTCTGGCATTCCTGGCATTCCTCCAGGTGCTGCTCCACATACATG 1396
 QY 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
 Db 1397 AAGAATCTGGACATAGTCGCCCA----- 1420
 QY 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
 Db 1421 -----GACAAAGAGACGACAG----- 1435
 QY 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
 Db 1436 -----AAGCAACAGGATTCACCTACCTATTAAAGATTGAAGTGGCAAGTGGG 1493
 QY 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
 Db 1484 ATGATACTGCTCATGGCTTTCTTCTCATCCTACCTTTCATTCAACCTGGGTGACC 1543
 QY 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
 Db 1544 AGTGAAGCTTACTCTTCTCCGTCCATTGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1600
 QY 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAlaAsp 549
 Db 1601 AGGATCATATTTGATGACTTCCGAGAGCATATATTTGGCTTCGTATATCTCCAGAG 1660
 QY 550 AspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
 Db 1661 GATCGAAGCTCATGCTCTGCTGGGATTTATGCTATCAGATTACAGTTACAGCTATGCGCAACCGA 1720
 QY 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
 Db 1721 ACAATTTTATGGACAAATAACACATGGAATAATACCATATTTCTCGAGTAGGGCAGCGCA 1780
 QY 590 MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
 Db 1781 ATGGCTGCTCCACAGAGAAAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1840
 QY 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 629
 Db 1841 CTGCTCATTTTGGAGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1900
 QY 630 MetValArgTleAla-----GluGlyGluHisProLysAspIleLysGluSer 645
 Db 1901 ATGGTCCGATTTGGAGGAGCAGATACAGGCAACAT-----ATCAAGGAGAT 1951
 QY 646 AspTyrPheThrAspArgGlyGluPheArgValAppAlaGluGlyAlaProAlaLeuLeu 665
 Db 1952 GACTATTATCTCAACTGGGAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2011
 QY 666 AsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLysLeuAspTyrArg 685
 Db 2012 AACTGCTCATGATACAGATGCTTACTATGCTTTGTCAGGTTTACACAGAGCCAG 2071
 QY 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu 705
 Db 2072 CGTCTCCAGGCTTGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2131
 QY 706 ThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLys 725
 Db 2132 GATGCTCTGGAGGAGCTATACACAGAACATTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2191

RESULT 10
 US-10-028-384-11
 ; Sequence 11 Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: FERRELL, Claude
 ; APPLICANT: MCBRIDE, Kevin
 ; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy


```

; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002219
; DATABASE ENTRY DATE: 2000-10-31
; RELEVANT RESIDUES: (1)..(2472)
US-10-028-384-11

```

Alignment Scores:		
Pred. No.:	7.22e-214	Length:
Score:	217.50	Matches:
Percent Similarity:	72.92%	Conservative:
Best Local Similarity:	57.54%	Mismatches:
Query Match:	53.92%	Indels:
DB:	15	Gaps:
		9
		2472

US-10-028-384-8 (1-774) x US-10-028-384-11 (1-2472)

QY	17	Ser	Leu	Leu	Thr	Phe	Ala	Ile	Leu	Leu	Ile	Leu	Leu	Ala	Trp	Leu	Ala	Gly	Phe	Ser	Arg	Leu	36
DB	152	AC	ACT	TTT	GGT	GA	AG	CTT	CTC	ATT	CTG	TC	AA	TGG	TG	TG	TAT	TAT	TAT	TAT	TAT	TAT	211
QY	37	Phe	Ala	Val	Ile	Asp	Glu	Ser	Ile	Ile	His	Glu	Phe	Asp	Pro	Trp	Phe	Asn	Tyr	Arg	56		
DB	212	TTT	GGT	CTC	CTG	GAG	ATT	TTG	AA	G	G	T	TAT	CC	T	G	AG	T	T	AT	T	AT	271
QY	57	Ala	Thr	Ala	Tyr	Met	Val	Gln	Asn	Gly	Trp	Tyr	Asn	Phe	Leu	Asn	Trp	Phe	Asp	Glu	Arg	76	
DB	272	ACT	ACC	AGG	TT	CTC	GGC	T	C	G	C	G	A	G	G	G	T	T	T	A	A	A	331
QY	77	Ala	Trp	Tyr	Pro	Leu	Gly	Arg	Glu	Val	Gly	Thr	Val	Tyr	Pro	Gly	Leu	Met	Ile	Thr	96		
DB	332	GCT	GGT	AC	CC	TTT	TGG	A	C	AG	A	T	C	AT	T	T	T	T	T	A	T	T	391
QY	97	Ser	Gly	Gly	Ile	His	Trp	Leu	Leu	His	Val	Leu	Asn	Ile	Pro	Val	His	Sile	Arg	Asp	Ile	116	
DB	392	TCT	GTG	CA	A	T	CT	T	ACC	AT	G	T	CT	CA	T	T	T	T	T	T	T	T	451
QY	117	Cys	Val	Phe	Leu	Ala	Pro	Ile	Phe	Ser	Gly	Leu	Thr	Ser	Ile	Ser	Thr	Tyr	Ileu	Leu	Thr	136	
DB	452	TGT	G	TTT	T	CTT	T	GG	C	CT	T	CT	T	CT	CC	T	CA	CT	CC	AC	G	T	511
QY	137	Lys	Glu	Leu	Trp	Ser	Ala	Gly	Ala	Gly	Leu	Phe	Ala	Ala	Ser	Phe	Ile	Ala	Ile	Val	Pro	156	
DB	512	AA	G	AG	CT	CA	AG	G	T	C	AG	G	G	G	CT	T	CT	T	T	G	CT	G	571
QY	157	Gly	Tyr	Ile	Ser	Arg	Ser	Val	Ala	Gly	Ser	Tyr	Asp	Asn	Glu	Gly	Ile	Ala	Ile	Phe	Ala	176	
DB	572	GGA	TAT	AT	CT	CCC	AT	CTG	GG	CG	T	CT	AT	G	A	T	A	A	G	A	G	A	631
QY	177	Leu	Gln	Phe	Thr	Tyr	Phe	Leu	Trp	Val	Arg	Ser	Val	Iys	Thr	Gly	Ser	Val	Phe	Trp	Ser	196	
DB	632	AT	G	T	ACT	CA	CT	T	ACT	T	GG	AT	CA	AG	C	AG	T	AA	G	AC	T	G	691
QY	197	Ala	Ala	Ala	Leu	Ser	Tyr	Phe	Tyr	Met	Val	Ser	Ala	Trp	Gly	Tyr	Val	Phe	Ile	216			
DB	592	GCT	AAG	T	GG	CC	CT	T	G	T	AT	T	CT	AT	C	AT	G	G	G	G	G	T	751
QY	217	Ile	Asn	Leu	Ile	Pro	Leu	His	Val	Phe	Val	Leu	Ile	Met	Gly	Arg	Tyr	Ser	Pro	Arg	236		
DB	752	AT	CA	ACT	T	T	A	T	CT	CT	CC	AC	G	T	CT	G	T	CT	CA	G	CC	G	811
QY	237	Leu	Leu	Thr	Ser	Tyr	Ser	Thr	Phe	Tyr	Ile	Leu	Gly	Leu	Leu	Phe	Ser	Met	Gln	Ile	Pro	256	
DB	812	AT	CT	AT	G	GG	CC	T	ACT	G	T	CT	AT	G	CT	GG	T	ACT	AT	CT	T	CT	871

Qy	257	PheValGlyPheGlnProIleArgThrSerGluHisMetalAlaLeuGlyValPheVal	276
Db	872	TTTGTGGGTTCCTCAGCCTGTCTCTTCATCAGACACATGGCAGGGTTGGGGTCTTTGGT	931
Qy	277	LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe	296
Db	932	CTCTGCCAGATCCATGCTTTGTGGATTACCTGCGCAGCAAGTTGAATCCACAACAATTT	991
Qy	297	ArgIysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal	311
Db	992	GAAGTTCTTTTCCGAGAGCGTCATCTCTCTGTAGCCTTTGTCTCTCTCCCGTGGGA--	1048
Qy	312	PheValAlaValValLeuThrMetLeuGlyValValAlaProTrrPrrSerGlyArgPhe	331
Db	1049	-----GCTCTCTCATGCTGACA-----GGAAAAATATCTCCCTCGACGGGGCGTTTC	1096
Qy	332	TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer	351
Db	1097	TACTCACTGCTGGATCCCTCTTATGCTAAGAACAAACATCCCATCATTTGCTTCTGTCTCT	1156
Qy	352	GluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla	371
Db	1157	GAGCATCAGCCCAACACTGCTCTCATATTATTGACCTGCAGCTCCCTCGTCTTCATG	1216
Qy	372	PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal	391
Db	1217	TTTCCAGTGGCCTCTATTACTGTTTGTAGCAACTGCTCATGCCCGGATTTTATCATC	1276
Qy	392	LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu	411
Db	1277	ATGTATGGTGTGACCAAGCTACTTTTTCAGCTGTAAATGGTGGGTCTTAATGCTAGTGTG	1336
Qy	412	ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu	431
Db	1337	GCACCTGTTATGACATTCTCTCTGGCATTTGGAGTCTCCAGGTGTGTCCACATACATG	1396
Qy	432	GlnGluAspSerSerIysArgMetGlyThrAlaIleSerAlaAlaThrGluValAlaSpGlu	451
Db	1397	AAGAACTGGACATAAGTCGCCCA-----GACAGAAGAGCAAG-----	1420
Qy	452	AlaGluAspSerIleGluIysLysThrLeuTyrAspIysAlaGlyLysLeuLysHisArg	471
Db	1421	-----GACAGAAGAGCAAG-----	1435
Qy	472	ThrLysHisAspAlaGlnAspThr-----GlyValSerSerAsnLeuLysSerIle	489
Db	1436	-----AAGCAACAGGATTCACCTACCTATTAAAGATTCAAGTGGCAAGTGGG	1483
Qy	490	ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr	509
Db	1484	ATGATACTGTGTAGCGTTCTTTCTCATCACCTACACCTTCATTCAACCTGGGTGACC	1543
Qy	510	SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer	529
Db	1544	AGTGAGCGCTACTCTTCTCGTCCATTGACTATCTGCCCGGTGGG---GATGGCAGT	1600
Qy	530	ArgAsnIleLeuAspAspPheArgGluAlaTyrTrrPrrLeuSerGlnAsnThrAlaAsp	549
Db	1601	AGGATCATATTGATGACTCCGAGAAGCATATTATTGGCTTCGTATAATACTCCAGAG	1660
Qy	550	AspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg	569
Db	1661	GATGCCAAGGTGATGCTCTGGTGGGATTATGGCTATCAGATTACAGCTATGGCAACCGA	1720
Qy	570	ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla	589
Db	1721	ACAATTTTATGTGCAATAACATCGGAATAATACCATATTATTCTCGAGTAGGGCAGGCA	1780
Qy	590	MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal	609
Db	1781	ATGGCGTCCACAGAGGAAAAACCTATGAGATCATAGGGAGCTCGATGTACGTATGTG	1840
Qy	610	LeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp	629

1841	DB	CTGTCATTTTGGAGCCCTCACTGGGTATCTCTGATGATATCAACAGTTCTTTGG	1900
630	QY	MetValArgIleala-----GluGlyGluHisProLysAspIleLysGluSer	645
1901	DB	ATGTCCTCGGATTGGAGGGACAGATACAGCAACAT-----ATCAAGGAGAA	1951
646	QY	AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu	665
1952	DB	GACTATTATACTCCAACTGGGGAGTTCCGTGTGGACCGTGAAGTTCTCCAGTCTCTCTC	2011
666	QY	AsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArg	685
2012	DB	AACTCGCTCATGTACAGATGTGTTACTATCGCTTTCACAGGTTTACACAGGCCAAG	2071
686	QY	GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu	705
2072	DB	CGTCCTCAGGCTTTACCGCTGTCGGAATGCTGAGATTGGGAATAAGACTTTTGAGCTT	2131
706	QY	ThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLys	725
2132	DB	GATGCTCTGGAGGAAGCTTATACCAGAACATTTGCTCGTCAGATATACAAGGTAAAG	2191

RESULT 11

```

RESUL. 11
US-10-172-118-742
; Sequence 742, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yucang
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172.118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM 002219
; DATABASE ENTRY DATE: 2001-06-18
; US-10-172-118-742

```

Alignment Scores:					
Pred. No.:	7,22e-214	Length:	2472		
Score:	2177.50	Matches:	415		
Percent Similarity:	72.9%	Conservative:	110		
Best Local Similarity:	75.6%	Mismatches:	144		
Query Match:	53.8%	Indels:	51		
DB:	15	Gaps:	9		
US-10-028-384-8 (1-774) x US-10-172-118-742 (1-2472)					
Qy	17	SerLeullethrPheAlaileLeulleuileAlaTrlpheLeuAlaGlyPheSerSerArgIeu	36		
	:	:::::	:::	:	:
Dd	152	ACACITTTGAAGCTTCTCATTCGTGCAATGCTGCTATTATTCCTCCACTCGICTG	211		
Qy	37	PheAlaVallllAArgPhecluSerllelehlsluPheAspProfrlpheAsnTyArg	56		
	:	:::::	:::	:	:
Dd	212	TTTGCTGTCTCAGATTTTGAAGAGTGTATTCATGAGTTTGCATCGTACITTAATTAICGG	271		
Qy	57	AlaThrAlataryMetValcInlAsnGlyTriptyrAsnPheLeuAsnTrlpheAspGluarq	76		
	:	:::::	:::	:	:
b	272	AATAACAGCTTCCTGCCTCAGAGAGGGTTTTATAAATTCATTAACITGGTTTTCATGACCGA	331		

77	AlaTrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThr	96
332	GCCTGGTACCCCTTTGGGACGAATCATTTGGAGGAACAATTTACCCAGGTTTAATGATCACC	391
97	SerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIle	116
392	TCTGCTGCAATCTACCATGTAATCTTCTTCCACATCACCATCGCAATTCGGAATGTC	451
117	CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr	136
452	TGTGTGTTCTCGGGCCCTCTCTCTCTCCCTTCACCTCCATCGTCAGCTACCTCCTTACC	511
137	LysGluLeuTrpSerAlaGlyValGlyLeuPheAlaAlaSerPheIleAlaIleValPro	156
512	AAAGAGCTCAAGGATGACGGGGCTTCTTCTGCTGCATCAATGCTGTAGTTCCT	571
157	GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla	176
572	GGATATATCTCCGATCTGTGGCTGGCTCTATGATATGAAGGATATGCCATCTTTGCG	631
177	LeuGlnPheThrTyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSer	196
632	ATGCTACTCACCTACTACATGTGGATCAAGGCAGTAAGACTTGGTTCCATCTGTGGGCA	691
197	AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIle	216
692	GCTAAGTGTGGCCTTGCTATTCTTACATGCTCGTCTCATGGGAGGTTATGTGTTCCTG	751
217	IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg	236
752	ATCAACTAAATTCCTCTCCAGCTCTCGTGTGATGCTCACAGGCGGTTCTCTCACCGG	811
237	LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro	256
812	ATCTATGTGGCCTACTGTACTGTTTTACTGCTCGGTACTACTTCTTAGCGAGATCTCC	871
257	PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal	276
872	TTTGTGGGTTCCAGCCTGTCTTTTCATCAGAGCACATGGCAGGGTTTGGGGTCTTTGGT	931
277	LeuLeuMetAlaValAlaThrLeuArgHisIleGlnSerValLeuSerArgAsnGluPhe	296
932	CTCTGCCAGATCCATGCCTTTGTGGATTACTCTCGCAGCAAGTTGAAATCCACAACAAT	991
297	ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal	311
992	GAAGTTCTTTTCGGAGCGTCATCTCTGTTAGGCTTGTCTCTCACCGTGGGA--	1048
312	PheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe	331
1049	-----GCTCTCTCATGTGACA-----GGAAAAATATCTCCCTGCAGCGGGCGTTTC	1096
332	TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer	351
1097	TACTTACTGCTGGATCCCTTTATGCTAAGAACACATCCCATCATTCCTTCTGTGTCT	1156
352	GluHisGlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla	371
1157	GAGCATCAGCCACACCTGGTCCTCATACTATTGTACCTCGACTCCTCTCTCTCATG	1216
372	PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal	391
1217	TTTCCAGTTGGGCTCTATTACTGCTTTAGCAACCTGCTCTGATGCCCGGATTTTATCATC	1276
392	LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu	411
1277	ATGTATGGTGTGACCAAGCATGTACTTTTTCAGCTGTAAATGTCGCTCTAATGCTAGTGTG	1336
412	ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu	431
1337	CGAGCTTCTTTCAGCAATTTCTTCTGATTCGATCTTCCAGTGTCTCCAGTGTCTGCCACATACATG	1396

QY 432 GlnGluAaspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAaspGlu 451
 Db 1397 AAGAATCTGGACATGAAGTCGCCA-----1420
 QY 452 AlaGluAaspSerIleGluLysThrLeuTyRaspLysAlaGlyLysLeuLysHisArg 471
 Db 1421 -----GACAAAGAAGCAG-----1435
 QY 472 ThrLysHisAspAlaGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
 Db 1436 -----AAGCAACAGGATCCACCTACCCTATTAAAGTTGAAGTGCGCAAGTGG 1483
 QY 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
 Db 1484 ATGATACTGGTCAAGCTCTTCTTCATCATCACCTTTCATCAACCTGGGAGACC 1543
 QY 510 SerAsnAlaTyRSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
 Db 1544 AGTGAGGCTACTCTCTCCGTCATTTGATCTATCTGCTCCGCTGGTGGG---GATGGCAGT 1600
 QY 530 ArgAsnIleLeuAaspPheArgGluAlaTyRTrpLeuSerGlnAsnThrAlaAap 549
 Db 1601 AGGATCATATTGATGATCTCCGAGAGCATATTATTGGCTTCGTATATATCTCCAGAG 1660
 QY 550 AspAlaArgValMetSerTrpAspTyRgLyThrGlnIleAlaGlyMetAlaAsnArg 569
 Db 1661 GATGGCAAGTCATGCTCTGCTGGATTATGGCTATCAGATTACAGTATGGCAACCGA 1720
 QY 570 ThrThrLeuValAspAsnThrTrpAsnSerHisIleAlaLeuValGlyLysAla 589
 Db 1721 ACAATTTAGTGGCAATAACACATGAATTAATACCATATTCTCAGTAGGCGAGGCA 1780
 QY 590 MetSerSerThrGluGluLysSerTyRgLyMetThrSerLeuAaspValAspTyRVal 609
 Db 1781 ATGGCGTCCACAGAGGAAAAGCGCTATGATCATGAGGAGCTCGATGTCACATATGG 1840
 QY 610 LeuValIlePheGlyGlyValIleGlyTyRSerClyAaspIleAsnLysPheLeuTrp 629
 Db 1841 CTGGTCATTTTGGAGGCTCCTGCGGTATTCCTCATGATATCAACAGATTCTTGG 1900
 QY 630 MetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluSer 645
 Db 1901 ATGGTCGGATGGAGGAGCACAGATACAGGCAACAT-----ATCAAGGAGANT 1951
 QY 646 AspTyRThrAspArgGlyGluPheArgValAspAlaGlnGlyAlaProAlaLeuLeu 665
 Db 1952 GACTATTATATCAACATGGGAGTTCGCTGGACCGTGAAGGTTCTCCAGTCTGCTC 2011
 QY 666 AsnCysLeuMetTyRLeuLysLeuSerTyRArgPheGlyGluLeuLysLeuAaspTyRArg 685
 Db 2012 AACTGCTCATGTACAAGATGTGTTACTATCGCTTTGGACAGGTTTACAGAGCCAG 2071
 QY 686 GlyProSerGlyTyRAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu 705
 Db 2072 CGTCTCCAGGCTTTGACCGGTCCGAATGCTGAGATTGGGAATAAGACTTTGAGCTT 2131
 QY 706 ThrTyRLeuGluGluAlaTyRThrGluHisTrpLeuValArgIleTyRArgValLys 725
 Db 2132 GATGTCCTGGAGGAAGGCTATPACCAAGACATTGGCTGCTCAGGATATACAGGTAAG 2191

RESULT 12

US-10-342-887-742
 ; Sequence 742, Application US/10342887
 ; Publication No. US20040058340A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dai, Hongyue
 ; APPLICANT: He, Yudong
 ; APPLICANT: Linsley, Peter S.
 ; APPLICANT: Mao, Mao
 ; APPLICANT: Roberts, Christopher J.
 ; APPLICANT: Van 't Veer, Laura Johanna
 ; APPLICANT: Van de Vijver, Marc J.
 ; APPLICANT: Bernards, Rene

; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
 ; FILE REFERENCE: 9301-188-999
 ; CURRENT APPLICATION NUMBER: US/10/342,887
 ; CURRENT FILING DATE: 2003-01-15
 ; PRIOR APPLICATION NUMBER: 60/298,918
 ; PRIOR FILING DATE: 2001-06-18
 ; PRIOR APPLICATION NUMBER: 60/380,710
 ; PRIOR FILING DATE: 2002-05-14
 ; PRIOR APPLICATION NUMBER: 10/172,118
 ; PRIOR FILING DATE: 2002-06-14
 ; NUMBER OF SEQ ID NOS: 2699
 ; SEQ ID NO 742
 ; LENGTH: 2472
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-342-887-742

Alignment Scores:
 Pred. No.: 7,22e-214 Length: 2472
 Scores: 2177.50 Matches: 415
 Percent Similarity: 72.92% Conservative: 110
 Best Local Similarity: 57.64% Mismatches: 144
 Query Match: 53.82% Indels: 51
 DB: 16 Gaps: 9

US-10-028-384-8 (1-774) x US-10-342-887-742 (1-2472)

QY 17 SerLeuLeuThrPheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeu 36
 Db 152 ACATTTTGAAGCTTCTCATTTCTGCAATGGCTGCTGTATATTCCTTCTCCACTCGTCG 211
 QY 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyRArg 56
 Db 212 TTTGCTGCTCGAGATTGAAGTGTATTCATGAGTTTGATCGGTACTTTAATATATCGG 271
 QY 57 AlaThrAlaTyRMetValGlnAsnGlyTrpTyRAsnPheLeuAsnTrpPheAspGluArg 76
 Db 272 ACTACCAAGTTCTTGGCTGAGGAGGGGTTTTATAATTCATAACATGCTTTCATGACCGA 331
 QY 77 AlaTrpTyRProLeuGlyArgIleValGlyGlyThrValTyRProGlyLeuMetIleThr 96
 Db 332 GCCTGGTACCTTTGGGACGAATCATTTGGAGGAACAATTTACCCAGGTTAATGATCACC 391
 QY 97 SerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIle 116
 Db 392 TCTGCTCAATATCACCATGATGATCTCCATTTTCCACATCACCATCGACATTCGGAATGTC 451
 QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyRLeuLeuThr 136
 Db 452 TGTGTGTTCTGGCCCT 511
 QY 137 LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
 Db 512 AAAGAGCTCAAGGATGCGAGGGCTGGGCTTCTTCTGCTGCCATGATGATGATGATGATGAT 571
 QY 157 GlyTyRLeuSerArgSerValAlaGlySerTyRAspAsnGluGlyIleAlaIlePheAla 176
 Db 572 GGATATATCTCCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 631
 QY 177 LeuGlnPheThrTyRLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
 Db 632 ATGCTACTCCTACTACATGATGGATCAAGGCGAGTAAGAGACTGGTTCATCTGTTGGGCA 691
 QY 197 AlaAlaAlaAlaLeuSerTyRTrpMetValSerAlaTrpGlyGlyTyRValPheIle 216
 Db 692 GCTAAGTGTGCTTGTCTTATTCTACATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751
 QY 217 IleAsnLeuLeuProLeuHisValPheValLeuLeuLeuMetGlyArgTyRSerProArg 236
 Db 752 ATCAACTTAATCTCTCCACGCTCTCGCTGCTGATGCTCAGGCGGCTTCTCTCACCGG 811
 QY 237 LeuLeuThrSerTyRThrPheTyRLeuGlyLeuLeuPheSerMetGlnIlePro 256

Db 812 ATCTATGTCCTACTGTACTGTTTACTGCTGGTACTATATCTTTCTAGGCAGATCTCC 871
Qy 257 PheValGlyPheGlnProIleAthrSerGluHisMetAlaAlaLeuValPheVal 276
Db 872 TTTGTGGTTTCAGCCTGCTCTTTCATCAGACCATGCGAGGTTTGGGGTCTTTGGT 931
Qy 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
Db 932 CTCTGCCAGATCCATGCTTTGTGGATTACCTGCGCAGCAAGTTGAATCCCAACAATTT 991
Qy 297 ArgLysLeuPhe 311
Db 992 GAAGTTCTTTTCGGAGCGTCATCTCTGTTAGGCTTTGCTCTCTCACCGTGGGA 1048
Qy 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProTrrSerGlyArgPhe 331
Db 1049 -----GCTCTCTCATGCTGACA-----GGAAAAATATCTCCCTGGACGGCGGTTC 1096
Qy 332 TyrSerLeuTrrPaspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
Db 1097 TACTCACTGCTGATCCCTTATGCTAAGAACACATCCCATCATCTCTCTGTCCT 1156
Qy 352 GluHisGlnProThrTrrPheSerPhePhePheAspLeuHisIleLeuValCysAla 371
Db 1157 GAGCATCAGCCCAACACCTGCTCTCATCTATCTTTTACCTTCGACCTCTCTCTTCATG 1216
Qy 372 PheProValGlyValTrrPyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
Db 1217 TTTCCAGTTGGCTCTTATCTGCTTTAGCAACCTGCTGATGCGCGGATTTTATCATC 1276
Qy 392 LeuTyrAlaIleSerAlaValTrrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
Db 1277 ATGTATGGTGTGACCATGCTCTCTGCGATGCGAGTCTCCAGTCTCTCCACATACATG 1336
Qy 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
Db 1337 GCACCTGTTATGACCATCTCTGCGATGCGAGTCTCCAGTCTCTCCACATACATG 1396
Qy 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
Db 1397 AAGATCTGACATAAGTCGCCA----- 1420
Qy 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
Db 1421 -----GACAAAGAGCAAG----- 1435
Qy 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
Db 1436 -----AAGCAACAGGATTCACCTACCTTATTAAGATTGAAGTGGCAAGTGGG 1483
Qy 490 ValIleLeuAlaValLeuMetLeuMetMetMetPheAlaValHisCysThrTrrValThr 509
Db 1484 ATGATACCTGCTGCTCTCTCTCATCACCTACCTACCTTCATTCATCAACCTGGGTGACC 1543
Qy 510 SerAsnNlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
Db 1544 AGTGAGCCCTACCTCTCTCGTCCATPTGACTATCTGCGCGTGGTGGG-----GATGGCAGT 1600
Qy 530 ArgAsnIleLeuAspAspPheArgGluAlaTrrTrrTrrLeuSerGlnAsnThrAlaAsp 549
Db 1601 AGGATCATATTTGATCACTTCGAGAGCATATATTGCTTCGTCATAAATCTCCAGAG 1660
Qy 550 AspAlaArgValMetSerTrrPaspTrrPaspTrrGlyTrrGlnIleAlaGlyMetAlaAsnArg 569
Db 1661 GATCGGAGGTGATGCTCTGCTGGGATTTATGCTATCATGATTCAGATTCAGCTATCGCAACCGA 1720
Qy 570 ThrThrLeuValAspAsnAsnThrTrrAsnAsnSerHisIleAlaLeuValGlyLysAla 589
Db 1721 ACAATTTTACTGACCAATAACACATGGAATAATACCATATTTCTCGAGTAGGCGACGA 1780
Qy 590 MetSerSerThrGlnGluLysSerTrrGluIleMetThrSerLeuAspValAspTrrVal 609
Db 1781 ATGCGCTCCACAGAGAAAGCCATATGATGATCATGAGGAGCTCGATGTCAGCTATGTG 1840

Qy 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrr 629
Db 1841 CTGGTCATTTTGGAGGCTCTCACTGGGTATTTCTCTCATGATATCAACAGTTTCTTTGG 1900
Qy 630 MetValArgIleAla-----GluGlyHisProLysAspIleLysGluSer 645
Db 1901 ATGGTCCCGATTTGGAGGAGGACAGATACAGGCAACAT-----ATCAAGGAGAAT 1951
Qy 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
Db 1952 GACTATTATCTCCAACTGGGGAGTTCGTTGGACCGTGAAGTTCTCCAGCTGCTC 2011
Qy 666 AnCysLeuMetTrrLysLeuSerTrrTrrArgPheGlyGluLeuLysLeuAspTrrArg 685
Db 2012 AACTGCTCATGTACAAGATGTGTACTGCTTTGGACAGGTTTACAGAGCAAG 2071
Qy 686 GlyProSerGlyTrrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu 705
Db 2072 CGTCTCCAGGCTTTGACCGTTCGGAATGCTGAGATTGGGAATAAAGACTTTGAGCTT 2131
Qy 706 ThrTrrLeuGluGluAlaTrrThrThrGluHisTrrPheValArgIleTrrValLys 725
Db 2132 GATGCTCGAGGAAGGCTATACACAGAACTTGGCTGCTCAGGATATACAGGTAAG 2191
RESULT 13
US-10-032-585-6323
; Sequence 6323, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6323
; LENGTH: 2256
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6323
Alignment Scores:
Pred. No.: 8,96e-214 Length: 2256
Score: 2176.00 Matches: 434
Percent Similarity: 70.16% Conservative: 109
Best Local Similarity: 56.07% Mismatches: 169
Query Match: 53.78% Indels: 62
Gaps: 15
US-10-028-384-8 (1-774) x US-10-032-585-6323 (1-2256)
Qy 11 LysValAlaGlyTrrSer---SerLeuIleThrPheAlaIleLeuLeuIleAlaTrrLeu 29
Db 34 AAAGTATTGGGATAGATGTTGAACCTGATTAGAGTTTTTATTAAGTTATTTATTTATA 93
Qy 30 -----AlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIle 45
Db 94 TCGATAGACGGTGCAGCTATTTCTTCGTTTATTTTCGTTGATTCGATTGGAAGTATT 153
Qy 46 IleHisGluPheAspProTrrPheAsnTrrArgAlaThrAlaTrrMetValGlnAsnGly 65
Db 154 ATTATGAATTCGATCTCTGTTCAATTTCCGACCAACCAATATTTAGTCACTCATTC 213
Qy 66 TrrTrrAsnPheLeuAsnTrrPheAspGluArgAlaTrrTrrPheLeuGlyArgIleVal 85
Db 214 TTTTATGAATTTTGAATTTGTTGATGATAGAACTTGGTACCATTTGGGAGAGTCACT 273
Qy 86 GlyGlyThrValTrrProGlyLeuMetIleThrSerGlyGlyIleHisTrrPheLeuHis 105

Db 274 GGTGGTACTTTATATCCGGTTTAAATGGTGACTTCAGTGCCATT--TGG-----CAT 324
 Qy 106 ValLeuAsn-----IleProValHisIleArgSpileCysValPheLeuAla 121
 Db 325 ATTTTACGTGATTGGTTGGCTTACCCGTTGATATTAGAAATATTGTGTGTTATTAGCA 384
 Qy 122 ProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyrSer 141
 Db 385 CCAGTTTCTCGGGATTAACTGCAATTGTTACTTATTTTGTACTAAAGAAATGAAGGAT 444
 Qy 142 AlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSerArg 161
 Db 445 TCTAGTCAGAGATTATGGCAGCTATATTATGGGATTGCCAGGTTATATTTCAGA 504
 Qy 162 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 181
 Db 505 TCAGTGCGTGGTCTTATGATATATGAAGCAATTGCCATTACTTTATTATATGGCAACATTT 564
 Qy 182 PheLeuTyrValArgSerValLysThrGlySerValPheThrPheSerAlaAlaLeu 201
 Db 565 TATTTCTGGGATTAAATCAATGAATAATGGGTTCAGTTTCTATGCCACATTGCAGCATTA 624
 Qy 202 SerTyrPheTyrMetValSerAlaTyrGlyTyrValPheIleIleAsnLeuIlePro 221
 Db 625 TTCTATTCTATATGTTAGTGTGGTGGTGGATATGTTTCAATACCAATTTGATTTCCA 684
 Qy 222 LeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSerTyr 241
 Db 685 TTACACGATATTTGCTTGTGTTTTCATGGGTGCTGTTATAATGCCAACTTACACTGCTTAT 744
 Qy 242 SerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPheGln 261
 Db 745 ACTACATGTTGCTTGGTACTTTGGCATCAATGCAGATTCCATTCGTTGGGTTTTTA 804
 Qy 262 ProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAlaVal 281
 Db 805 CCAATTAGATCAATCATCATATGCTGCTGATTAGGAGTATTGGATTGTTACAAATTAGTG 864
 Qy 282 AlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPheIle 301
 Db 865 GCTTTTGGTGAATTTATGTTTAAATCAAAAGTTTCCAAACCAAAATTTAAATTCATCTTGATA 924
 Qy 302 ValGly---GlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 Db 925 GTTTCATGTTACTTGTGTTGGATTAGGTATT--GGTGGATTATTGGATTAAACACA 981
 Qy 321 LeuGlyValValAlaProTyrSerGlyArgPheTyrSerLeuTyrAspThrGlyTyrAla 340
 Db 982 ATGGGTGGATTGCTCTTGGACAGGTAGATTATTTCTTATGGATACAAATTTATGCC 1041
 Qy 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrPheSer 360
 Db 1042 AGAATTCATATTCCAATATGCTTCTGTTCTGACATCAACCTACTGCTTGGCCAGCA 1101
 Qy 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrTyrCysIle 380
 Db 1102 TTCCTTTTCGATAGTATGTTTATTTGGTTATTCCCGCTGGTGTATCTATTATGTTTC 1161
 Qy 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 Db 1162 CAAGAATTGAAGATGAACACGTTTTCATTTATCATTTACAGTGTATTGTGTTCTATTATT 1221
 Qy 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 Db 1222 GCTGGTGTATGCTGAAGATTGATTGTTGATTGCTTCCAGTCATTTGTTGTTGCTGAGCA 1281
 Qy 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgValGly 440
 Db 1282 ATTGCTTATCAATTAATTTGTTGATGCTATTG-----1314
 Qy 441 ThrAlaIleSerAlaAlaThrGluValAlaAspGluAlaGluAspSerIleGluLysThr 460

Db 1315 -----GACATTGTTGATTGTTCACT 1335
 Qy 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnIleAspThr 480
 Db 1336 -----GAGAAAGTTGAAAGTATGATGATGACGTTAGTGAATCCAAGAGCTCAACC 1389
 Qy 481 GlyValSerSerAsn-----LeuLysSerIleValIleLeu 492
 Db 1390 AAAAAATCAAGTTCCAGATTCCCAATTCTCGATATTTGTCAAAAGTTTGTGTTTACTG 1449
 Qy 493 AlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTyrValThrSerAsnAla 512
 Db 1450 ACAATTATACATTTTACCTTTTCTACTTTTACATGTTACTTGGTAACATCGAATGCT 1509
 Qy 513 TyrSerSerProSerIleValLeuAlaPheHisIleSerGlnAspGlySerArgAsnIle 532
 Db 1510 TATTCATCACCATTGATGTTTGTAGTCCAGAAC---CCAGATGGCTCCACACATATC 1566
 Qy 533 LeuAspAspPheArgGluAlaTyrTyrTyrLeuSerGlnAsnThrAlaAspAspAlaArg 552
 Db 1567 ATTGATGATTATAGAGAGCCTATTACTGTTTAAAGATGAATACACAGAAGATGCCAAA 1626
 Qy 553 ValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 572
 Db 1627 GTTATGGCTCGTGGGATTATGGTTATCAATCCGGGGTATGGCTGATAGAACACACTT 1686
 Qy 573 ValAspAsnAsnThrTyrAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 592
 Db 1687 GTTGATAACATATCATGGAATACACATATTTGCCACTGTTGGTAAGGCAATGCTCTCC 1746
 Qy 593 ThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIle 612
 Db 1747 CCTGAAGATGTCGTATGAAATTTTGAGACAACACACATGTTGATTATGTAGTTATA 1806
 Qy 613 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetValArg 632
 Db 1807 TTTGAGGGTTATTGGGTTATTCTGGTGATGATATTAAACAATTCCTTATGGATGGTAAGA 1866
 Qy 633 IleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPheThrAspArgGly 652
 Db 1867 ATTGCTGAAGTATCTGGCTGATGAATCAAGAAAGAGACTACTTTACTGACCGAGGA 1926
 Qy 653 GluPheArgValAlaAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMetTyrLysLeu 672
 Db 1927 GAATATAAGTGGATAAAGATGCATCTCGCAATGAAGAAATTCCTTGTATGATAAGTTA 1986
 Qy 673 SerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGlyTyrAspArg 692
 Db 1987 TCGTATCATAGATTCACTGATTG-----TTTGGAGGTAGAGATGTTGTTGATAGA 2037
 Qy 693 ThrArgAsnAlaValIle---GlyAsnLysAspPheAspLeuThrTyrLeuGluGluAla 711
 Db 2038 GTTAGAAACCAACAAATCCAGCCAATGAAGTACCGAAATTTGAATGTTTGAAGAAGCC 2097
 Qy 712 TyrThrThrGluHisTyrLeuValArgIleTyrArgValLysLysProHisGluPheAsn 731
 Db 2098 TTCATCATCAAAATTTGGATTGAGATTTACAAAGTTTAA-----2139
 Qy 732 ArgProSerLeuLysThrLysGluArgThrIleProProAlaAsnPheIleSerArgLys 751
 Db 2140 -----GATTGGATAATGTTGGTAGAGATTACATCAAGCTACTGCTTTTGAAGAATCA 2193
 Qy 752 AsnSerLysArgLysGlyTyrIleArgAsnArgProVal 765
 Db 2194 TCATCCGCACTTCCAAA-----AGAAACAGATCCATA 2236

RESULT 14
 US-10-128-714-7139
 ; Sequence 7139, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Wengli

APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIORITY APPLICATION NUMBER: US 60/285,697
PRIORITY FILING DATE: 2001-04-23
PRIORITY APPLICATION NUMBER: US 60/287,066
PRIORITY FILING DATE: 2001-04-27
PRIORITY APPLICATION NUMBER: US 60/295,890
PRIORITY FILING DATE: 2001-06-05
PRIORITY APPLICATION NUMBER: US 60/303,899
PRIORITY FILING DATE: 2001-07-09
PRIORITY APPLICATION NUMBER: US 60/316,362
PRIORITY FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: Patent version 3.1
SEQ ID NO 7139
LENGTH: 2232
TYPE: DNA
ORGANISM: Aspergillus fumigatus
US-10-128-714-7139

Alignment Scores:
Pred. No.: 2,036-213 Length: 2232
Score: 2172.50 Matches: 427
Percent Similarity: 71.24% Conservativity: 108
Best Local Similarity: 56.86% Mismatches: 171
Query Match: 53.76% Indels: 45
DB: 15 Gaps: 11

US-10-028-384-8 (1-774) x US-10-128-714-7139 (1-2232)

QY 18 LeuileThrPheAlaileLeuLeuileAlaTrpLeuAlaGlyPheSerSerArgLeuPhe 37
DB 61 CTTCTCGGATTATTATCTGCTGTACCATTCGACGAGCAGCAGTTCGCCAGTCGACTCTTC 120
QY 38 AlaValileArgPheGluSerleileHisGluPheAspProTrpPheAsnTrpArgAla 57
DB 121 AGCGTTATCGGCTTCGAGAGTATCATCCAGGAAGTTCACCGGTGTTCACTCCGAGCA 180
QY 58 ThrAlaTrpMetValGlnAsnGlyTrpTrpAsnPheLeuAsnTrpPheAspGluArgAla 77
DB 181 ACAATACTTAGTACAGATGGTTCTATAGCTTTTGGATTGGTTTGGATGACGGAACA 240
QY 78 TrpTrpProLeuGlyArgileValGlyThrValTrpProGlyLeuMetileThrSer 97
DB 241 TGGCATCTCTGGAGCTGTACCGGTGACGATATATCCGGTCTCATGGTACGAGC 300
QY 98 GlyGlyileHisTrpLeuHisValLeuAsnileProValHisileArgAspilleCys 117
DB 301 GCGGTGATCTACATATCTGGATCTCTTACTATCCCGCTCGATATTCGCAACATCTGC 360
QY 118 ValPheLeuAlaProPheSerGlyLeuThrSerileSerThrLeuLeuThrlys 137
DB 361 GTCTACTCGGCGCAGGATCTCCGCGCTGACTGCAATGGCAATGCTGTCATCTTC 420
QY 138 GluLeuTrp--SerAlaGlyAlaGlyLeuPheAlaAlaSerPheileAlaileValPro 156
DB 421 GAGATGTCCTTCGGCATCTGAGGTCTCTTGTGAGCAGCTTTTATGGGATTCGCCCT 480
QY 157 GlyTrpileserArgSerValAlaGlySerTrpAspAsnGluGlyileAlailePheAla 176
DB 481 GGTTACATCTCCCGATCAGTGTGGAAGTACGATACGAAAGCAGATTCGCACTCTTC 540
QY 177 LeuGlnPheThrTrpPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
DB 541 CTTGTGTTCATCTTCTTATGGATCAAGGCTGTCAAAATGGGTCTATCATGTGGGA 600

QY 197 AlaAlaAlaLeuSerTrpPheTrpMetValSerAlaTrpGlyTrpValPheile 216
DB 601 CCGCTGACCGGACTATTCTACGGCTACATGGTGTGGCATGGGTGGTATGCTTCATT 660
QY 217 IleAsnLeuileProLeuHisValPheValLeuLeuileMetGlyArgTrpSerProArg 236
DB 661 ACGAACCTGATCCCGCTGCACGTTTTTGTCTTCTGTGATGGGTAGATACAGACCTCGC 720
QY 237 LeuLeuThrSerTrpPheTrpPheTrpPheTrpPheTrpPheTrpPheTrpPheTrp 256
DB 721 ATCTACATTAGCTATACCATGATGCTGCGTGGGACTTTGGCTACATGCGATTCCTCC 780
QY 257 PheValGlyPheGlnProileArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
DB 781 TTCGTGCGATTTTTCCTTATCCGAAACAGCAGCAGCATGTCCGCTTGGGTGTCTTCGGC 840
QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
DB 841 CTGCTTCAGCTTGTGGCCTTCGCGAGTTTGTCCGAGCTTCGTTCCAAAGCAGCAGTTC 900
QY 297 ArgLysLeu-----PheileValGlyGlyLeu-----LeuValGlyValGly 310
DB 901 CAGACACTTCTGACCGCCATGATCTTCATCACCCTCGGTCTCGGTTCGTTGGTGG- 954
QY 311 ValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArg 330
DB 955 -----CTAGTTGTTCTGACTGTGACGGAGTATCGCTCTCTTGGAGCGGCGCA 1002
QY 331 PheTrpSerLeuTrpAspThrGlyTrpAlaLysileHisileProileAlaSerVal 350
DB 1003 TTCTACTCTTGTGGGACACTGGCTATGCCAAATCCACATTCCTCATCTTCGCTCAGTC 1062
QY 351 SerGluHisGlnProTrpThrTrpPheSerPhePhePhePhePhePhePhePhePhePhe 370
DB 1063 TCGGAACACAGCCACCGCTTGGCCAGCTTCTTCTTCGATCGAATCTCTGATCTCGG 1122
QY 371 AlaPheProValGlyValTrpTrpCysileLysGlnileAsnAspGluArgValPheVal 390
DB 1123 CTTTTCCCGGAGGTGTACATGTCTCGGTGACCTCAAGGAGCAGCATGTCTTCGTC 1182
QY 391 ValLeuTrpAlaileSerAlaValTrpPheAlaGlyValMetValArgLeuMetileThr 410
DB 1183 ATTATCTACTCGTCTTCGAGCTACTTCGCGGTGTATGTCGCGCTATGTCGACCTAT 1242
QY 411 LeuThrProValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPhe 430
DB 1243 TTGACCCCTATTGTGTGTCTCCGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1302
QY 431 LeuGlnGluAspSerSerLysArgMetGlyThrAlaileSerAlaAlaThrGluValAsp 450
DB 1303 ATGGCGACT-----ACCTCCCGACACACGAGCTCTGAGCGGAA 1341
QY 451 GluAlaGluAspSerileGlyLysThrLeuTrpAspLysAlaGlyLysLeuLysHis 470
DB 1342 ACGAATGAGAGACTCGTTCACAACT----- 1368
QY 471 ArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSerileVal 490
DB 1369 ---CTTGGCTCAGTTCGGAAGCCCAATGTTGAATCACCTCCCATGTTTCTTAAGATTATA 1425
QY 491 IleLeuAla---ValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThr 509
DB 1426 GTGACGCGCTCTCTGCTCTACCTGCTCTCTGTTGTTGCGCAGCTGACCTGGGTATCA 1485
QY 510 SerAsnAlaTrpSerSerProSerileValLeuAlaPheHisAsnSerGlnAspGlySer 529
DB 1486 TCGAATGATACACTCTCTCTCTCGTGTCTTGGCT---AGTCGGATGCTGACGGAAGC 1542
QY 530 ArgAsnileLeuAspAspPheArgGluAlaTrpTrpLeuSerGlnAsnThrAlaAsp 549
DB 1543 CAATACATCATTCAGCATTCATCGTAGGCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1602
QY 550 AspAlaArgValMetSerTrpTrpAspTrpGlyTrpGlnileAlaGlyMetAlaAsnArg 569

QY 361 PhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrCysile 380
DB 1258 TTCTACTTTGACCTCGAAGGCTTAATCTTTTCCCTGCGGTGCTTCTGGGTTC 1317
QY 381 LysGlnIleAsnAspGluArgValPheValLeuTyrAlaIleSerAlaValTyrPhe 400
DB 1318 AAGGAGCTTCGGATGACGAGATCTTCATCATCATTTATGCGGTCTCAGTGCCTATTATTT 1377
QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
DB 1378 GCCGCTGTCATGTTGCTGACTATGCTGTATCATCACCGCTGTGTCTGTGTTCTCCGCC 1437
QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
DB 1438 ATTGCGTTCTCCAAACTTCTCGAGCGGTATATTGAC----- 1473
QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
DB 1474 -----CCGTCATCCCGAAGCGAC-----GAGGAAGCTGGGAGTCTCAGACG 1518
QY 461 LeuTyrAspLysAlaGlyLysLysHisArgThrLysHisAspAlaGlnAsp--- 479
DB 1519 CAGGTTGCTCCAAAGTCCAGGCGAAGAAGATGGCGCTGCCAACGCCCAATAAGAGCGGG 1578
QY 480 -----ThrGlyValSerSerAsnLeuLysSerIle----- 489
DB 1579 TTCTCTTTACAGGTATTTGACGGC---NAGTCTGTCTCGGCATCTTTGGTCTCGAC 1635
QY 490 -----ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCys 505
DB 1636 ACTCGATTGCTGTGGTTTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1695
QY 506 ThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSer 525
DB 1696 ACATATGTGACTTCACAGCGGTATTTCTGCGCTTCAGTGGTACTTGCATCGCGAAC--- 1752
QY 526 GlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGln 545
DB 1753 CCGGATGGTAGCCAAATATCATTTGATGATTTCCGAGAGGCTTACTACTGGATTGCGCAA 1812
QY 546 AsnThrAlaAspAlaArgValMetSerTrpTyrTrpAspTyrGlyTyrGlnIleAlaGly 565
DB 1813 AACACCCCGGACAGACGCTCATCATGCTCTGGTGGATTACGGCTACAGATCGCTGT 1872
QY 566 MetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeu 585
DB 1873 ATGGCTGATCGCCCGCCCTTTGTTGATAACAATACCTGGAAATACACCCACATTCGCCACA 1932
QY 586 ValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAsp 605
DB 1933 GTTGGTAAGGCCATGGCTTCCACGAGATGTCGATATCTCTCTCTGAGGAGCATGAT 1992
QY 606 ValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspIleAsn 625
DB 1993 GTCGATTACGTTCTTTGATCTTTGGGGCTTATTGGGCTACTCTGGTGACGATATCAAC 2052
QY 626 LysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSer 645
DB 2053 AAGTTTGTGGATGGTTAGGATCTCACAGGTGATGGCTCTGACAGGTGCGAGNAGTC 2112
QY 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
DB 2113 AACTACTTTTACTCAAGAGGGGAGTATGCTGTCGATGACAGGGGCCACCCCTTACTATGAAG 2172
QY 666 AsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLysAspTyrArg 685
DB 2173 AACTCTCTCATGTACAAATGTCTTACTACCGCTTCCCGAGCTT-----TATGGT 2223
QY 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu 705
DB 2224 GGACACCCCGCTCAAGACAGGGTTGAGGCCAAATTTATCCCTCTCTCAACAGTGTACTCTT 2283

QY 706 ThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLys 725
DB 2284 GATACTCTTTGACGAGCGTTTCACATCCGAAAATTGGATCGTCAGGATCTACAAGGTCAAG 2343
QY 726 LysProHisGluPheAsnArg 732
DB 2344 AAGGAAGATCCCATTTGGACGA 2364

Search completed: December 15, 2004, 21:18:32
Job time: 1112.48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:52:55 ; Search time 5991.78 Seconds
(without alignments)
4707.175 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 4046

Sequence: 1 MNRTPKMLNSKVAGYSSLIIT.....RRKGYIRNPVVVKGKRTLK 774

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-NODEL=frame+...p2n.model -DEV=xlh
-Q/cgn2_1/USPTO_epool/US10028384/runat_14122004_131519_8030/app_query.fasta_1.3740
-DB=EST -QFWT=fastap -SUFFIX=1rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=pto -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN 1.13960 @runat_14122004_131519_8030 -NCPU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hsc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2948	72.9	2709	3 AK018758	AK018758 Mus muscu
2	2947.5	72.8	4513	3 BC044321	BC044321 Xenopus l
3	2906	71.8	2656	3 AK012153	AK012153 Mus muscu
4	2200	54.4	2656	3 BC028897	BC028897 Mus muscu
5	2191.5	54.2	3802	3 AK030363	AK030363 Mus muscu
6	2185.5	54.0	2730	3 AK081547	AK081547 Mus muscu
7	2170.5	53.6	3643	3 AK077877	AK077877 Mus muscu
8	2071.5	51.2	2118	9 AY418284	AY418284 Homo sapi
9	2067	51.1	2727	3 AK087470	AK087470 Mus muscu

10	2053.5	50.8	2334	3	AY310157	AY310157 Rattus no
11	2044	50.5	2103	9	AY418286	AY418286 Mus muscu
12	2036.5	50.3	2419	3	CNS0AD12	BN16490 Arabidops
13	1669.5	41.3	2079	9	AY418285	AY418285 Pan trogl
14	1396	34.5	836	9	CNS0172X	AL108711 Drosophi
15	1331	32.9	810	2	BF502026	BF502026 AT17637.5
16	1196.5	29.6	960	6	CD793779	CD793779 EST665140
17	1190	29.4	694	4	BI354111	BI354111 GM26338.5
18	1187	29.3	876	6	CD794446	CD794446 EST665807
19	1176	29.1	701	2	BF503420	BF503420 AT19406.5
20	1174	29.0	878	6	CD788585	CD788585 EST659946
21	1167	28.8	891	6	CD778215	CD778215 EST649576
22	1155	28.5	852	7	CN458732	CN458732 UI-M-HBO
23	1149.5	28.4	1345	3	CR606556	CR606556 full-length
24	1149	28.4	715	7	CK657451	CK657451 LP23759.5
25	1143.5	28.3	919	7	CK178331	CK178331 EST767651
26	1143	28.3	736	1	AA949890	AA949890 LD29946.5
27	1136	28.1	685	2	BF496296	BF496296 AT10060.5
28	1125	27.8	690	4	EG640860	EG640860 SD12024.5
29	1122.5	27.7	950	5	BU515954	BU515954 AGENCOURT
30	1116	27.6	2212	3	CR610915	CR610915 full-length
31	1103	27.3	823	7	CF736482	CF736482 UI-M-HDO
32	1098	27.1	679	7	CK657649	CK657649 LP24119.5
33	1098	27.1	681	7	CK659033	CK659033 LP16462.5
34	1083	26.8	682	7	CK659064	CK659064 LP16523.5
35	1078	26.6	978	7	CR413615	CR413615 CR413615
36	1053.5	26.0	1145	5	EX446691	EX446691 EX446691
37	1052	26.0	896	6	CA980464	CA980464 AGENCOURT
38	1051	26.0	765	6	CA382416	CA382416 662214 NC
39	1046.5	25.9	784	6	CD784240	CD784240 EST655601
40	1045	25.8	652	4	BI171940	BI171940 RE13841.5
41	1043	25.8	798	7	CN527592	CN527592 UI-M-HQO
42	1042	25.8	663	4	EG641172	EG641172 SD12448.5
43	1041	25.7	666	4	BI357074	BI357074 RE43425.5
44	1041	25.7	669	4	BI227902	BI227902 RE25388.5
45	1037	25.6	641	1	AI257750	AI257750 LP06212.5

ALIGNMENTS

RESULT 1
AK018758
LOCUS
DEFINITION
Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:130006C19 product:hypothetical Oligosaccharyl transferase (OraSe) SRT3 subunit containing protein, full insert sequence.
AK018758
2709 bp mRNA linear HTC 03-APR-2004
AK018758.2 GI:26384577
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
2
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, T., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 MEDLINE
 PUBMED
 11076861
 4
 AUTHORS
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.

TITLE
 JOURNAL
 REFERENCE
 5
 AUTHORS
 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.

TITLE
 JOURNAL
 REFERENCE
 6
 AUTHORS
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

TITLE
 JOURNAL
 REFERENCE
 6
 AUTHORS
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
 Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
 Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
 Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Konda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
 Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
 Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
 Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
 Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
 Tejima, Y., Toyota, Y., Yamanura, T., Yasunishi, A., Yoshida, K.,
 Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission

TITLE
 JOURNAL
 REFERENCE
 6
 AUTHORS
 Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp).
 URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216

COMMENT
 On Dec 10, 2002 this sequence version replaced gi:12858635.
 Please visit our web site (<http://genome.gsc.riken.jp/>) for further
 details.

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5'-GAGAGAGAGCGGCGCCACTCGAGTTTCTTTTCTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. Second
 strand cDNA was prepared with the primer adapter of sequence [5'-
 GAGAGAGAGAGCGGCGCCACTCGAGTTTCTTTTCTTNN 3']. cDNA was
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
 XhoI. Host: SOUR.

FEATURES
 source
 1..2709
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:1300006C19"
 /db_xref="taxon:10090"
 /clone="1300006C19"
 /sex="male"
 /tissue_type="liver"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 76..2547
 /note="unnamed protein product; hypothetical
 Oligosaccharyl transferase (OTase) SPT3 subunit containing
 protein (InterPro|IPR003674, evidence: InterPro)
 putative"
 /codon_start=1

CDS

/protein_id="BAB31390.1"
 /db_xref="GI:12858636"
 /translation="MASEPAPESKHSLSNPSWGLMAGNRHGHGPGTQSASRA
 AAEFGPAGLSGGLSPAGWQSLSTFLFLAWLAGFSRLFAVIRFESIHEPDPW
 FNYRSTHLAGHGFELFNWDERAWYFLGRIVGTVYPLMTAGLILWINTLNTIT
 VHIRDVCFLAPTFSLTSTFLTRELNQAGLILACFIAIVPGYISRSVAGSFD
 NEGDAIPALOPTYVLYWVSKTSVFTWCCCLSYFVMSVAMGVYFINLIPLVHFV
 LLMORYSKRVYIAVSTLYLVGLLSMQIPFVGQPIRTSEHMAAGVALLQAVFL
 CYLDRLLTKQEFQILFLGVSILACAVLSVILTYTGYIAPWSGRPYSLMTGYAKI
 HIPIIASVSEHQPTWVSFFDLHLVCTFPAGLNFCKINRDERVFVLYALISAVYF
 AGVVRMLTLTPVPCMLSAFANFNVFHYGLMDMKRNPVEDSDKKNPGLV
 DKAGVRKHVTQEPKEGLGNPKISIVTLMMLLMFMFAVHTWTISNAYSSPVLY
 ASYNDGTNRLDDREFAYFLWRQNTDHAHVMSWMDYGYAGANRITLVNDNTN
 NSHALLVGKMSNSTEAYKIMRSLDVLVDVYLVIFGVGVIGYGGDDINKFLVMVRIAGE
 HPKDIRGDYFTQGEFRVDKAGSPVTLNCLMKYSYRFGEMQLDFRTPPGQDRTRN
 AEIGNKDIKPKHLEASTSEHVALVRIYKVKAPDNRETLGHKPRVTNIVPKQVLSKKT
 TKRXRGVYKVKLVFKKTKSKTV"
 2693..2698
 /note="putative"
 2709
 /note="putative"
 /polyA_signal
 /polyA_site

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2709
 Score: 2948.00 Matches: 551
 Percent Similarity: 83.14% Conservativeness: 90
 Best Local Similarity: 71.47% Mismatches: 114
 Query Match: 72.86% Indels: 16
 DB: 3 Gaps: 5

US-10-028-384-8 (1-774) x AK018758 (1-2709)

QY 10 SerLysValAlaGlyTyrSerSerLeuLeuThrPheAlaLeuLeuLeuLeuLeuLeu 29
 DB 253 TCGCAGCGCGCGCGGTGGCGAGTCGTCCTCTCCATCCATCCTCTCTCTGGCGTGG 312
 QY 30 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 49
 DB 313 GCCGCGTTCAGTCGCGCCCTTCGCGCGTCATCGCTTCGAGAGCATCATCCACGAGTTC 372
 QY 50 AspProThrPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPhe 69
 DB 373 GACCGCGTGTAACTATAGATCAACACATCTTCATCTCATGATCTCATGATGATTT 432
 QY 70 LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyGlyThrVal 89
 DB 433 CTAAATTCGTTTGTATGAAGAGCATGTATCCCATCGGAGAGATAGTGGGTGGCACCGTT 492
 QY 90 TyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIle 109
 DB 493 TACCCAGCGTGTATGAACAGCTGGCGCTTATTCATTTGATTTTAAATCAATGAAACATA 552
 QY 110 ProValHisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer 129
 DB 553 ACAGTTCACATAAGAGATGTGTGTATTCCTTGCACCACTTTTAGCGGCTTATCATCC 612
 QY 130 IleSerThrTyrLeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAla 149
 DB 613 ATATCTACGTTCTCTGCTAACTAGAGAACTGTGAACCAAGGAGCAGACTTCTAGCTGCT 672
 QY 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 169
 DB 673 TGCITTCATTCATCGTACCGAGGTATATCTCGGTTCAGTGGCGGATCTCTTGATAT 732
 QY 170 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLys 189
 DB 733 GAAGCATTGCGATTTTGGCGTTCAGTTCACTTACTTATGGGTAAAGTCTGGAAG 792
 QY 190 ThrGlySerValPheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAla 209
 DB 793 ACCGGGTCTGTGTCGACCAATGTGCTGCTGTGCTGTGCTATTTTACATGCTCTCTGCG 852

Qy 590 MetSerSerThrGluGluLysSerTyrGluLeuMetThrSerLeuAspValAspTyrVal 609
 Db 2161 ATGCTTCAATGAACCGCTGTTATGAAATATGAAAGCTTGATGATATGATGA 2220
 Qy 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 629
 Db 2221 CTAATAATATTGGGGGTGTAATTGGATATCTGGTGATGACATCAACAAATCTCTCGG 2280
 Qy 630 MetValArgIleAlaGluGlyGluHisProLysAspIleLeuGluSerAspTyrPheThr 649
 Db 2281 ATGTTTCCAATAGCAGAGAGAGAACATCCAAAGATATCAGGAAAGTCACTACTCACC 2340
 Qy 650 AspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMet 669
 Db 2341 A--CAAGGAGAGTTTCGTGTAGACAAAGCTGGATCTCCAACTTGTCTCAATTCCTCATG 2398
 Qy 670 TyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGly 689
 Db 2399 TATAGATGTCATATATCCGGTGTGGGAAATCGACGCTGATTTCCGAAACACACCTGGA 2458
 Qy 690 TyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeuGlu 709
 Db 2459 TTTGACCGGACACGCTAATGCTGAAATGGGAATAGGATATTAATTCAGCATTTGGAA 2518
 Qy 710 GluAlaTyrThrThrGluHisTyrLeuValArgIleTyrArgValLysLysProHisGlu 729
 Db 2519 GAACATTTACATCGGAGCACTGGCTGTGTAGATATACAAAGTTAAAGAGCTAGAA--- 2575
 Qy 730 PheAsnArgProSerLeuLysTyrLysGluArgThr-----IleProProAlaAsn 746
 Db 2576 --AACAGAGAGGCACCTTGATCATATAACCTAGATTAACCAATATTGTACCAAAACAGAA 2632
 Qy 747 PheIleSerArgLysAsnSerLysArgGlyGlyTyrIleArgAsnArgProValVal 766
 Db 2633 TATTTATCAAAAAGACATCTAGAGAGAGCGTGCTACATTAGGAACAGCTGACACATG 2692
 Qy 767 ValLysGlyLysArgThrLeuLys 774
 Db 2693 AAGAAAGGCAAGAGACCAACAAAG 2716

RESULT 3
 LOCUS AK012153 2669 bp mRNA linear HTC 03-APR-2004
 DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
 enriched library, clone:2610524N02 product:hypothetical
 Oligosaccharyl transferase (OTase) STT3 subunit containing protein,
 full insert sequence.

ACCESSION AK012153
 VERSION AK012153.1 GI:12848721
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Chara, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 TITLE The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 JOURNAL Functional annotation of a full-length mouse cDNA collection
 REFERENCE Nature 409, 685-690 (2001)
 5
 TITLE The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 JOURNAL Analysis of the mouse transcriptome based on functional annotation
 REFERENCE of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 2669)
 AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
 Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
 Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
 Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Mishi, K.,
 Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
 Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
 Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
 Suzuki, H., Tashiro, M., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
 Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
 Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 Please visit our web site (http://genome.gsc.riken.jp/) for further
 details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5'-GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. Second
 strand cDNA was prepared with the primer adapter of sequence [5'-
 GAGAGAGATCTCGAGTTAATTAATTAATCCCCCCCCC 3']. cDNA was cleaved
 with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
 Host: SOLR.

FEATURES
 Location/Qualifiers
 1..2669
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:2610524N02"
 /db_xref="taxon:10090"
 /clone="2610524N02"
 /tissue_type="whole body"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="10 days embryo"
 misc_feature 36..2507
 /note="hypothetical Oligosaccharyl transferase (OTase)
 STT3 subunit containing protein (InterPro|IPR003674,
 evidence: InterPro)
 putative"
 polyA_signal 2654..2659
 /note="putative"
 polyA_site 2669

Db 2168 CTGCTTATGATAAAATGTCATCTACAGATTGGAGAAATCGACGTAGATTTCGCAC 2227
 QY 686 yProSerGlyThrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuth 706
 Db 2228 TCCCCAGGCTTACCGAACACAGTAATGTCAGATTGGAATAAAGACATTAAATTCAA 2287
 QY 706 rTyrlcugluclualatyThrThrGluHisTrrLeuValarglletyThrGlyValysly 726
 Db 2288 GATTITGGAGAAGCTTTTATCATCAGACACTGCGCTTGTGAGATATATAGGTAAGC 2347
 QY 726 sProHisGluPheAsnArgProSerLeuLysThrLysGluArgThr-----llePr 743
 Db 2348 ACCTGAC-----AACAGGAGACACTAGGTGCACAAACCTCGAGTCACCAACATCGTCC 2401
 QY 743 oProAlaAsnPhelSerArgLysAsnSerLysArgArgLysGlyTyrlleArgAsnAr 763
 Db 2402 CAACAGAGATATTGTCAAGAGAGACTACTAAAGGAGCGTGGCTTACGTTAAATAAT 2461
 QY 763 gProValValLysGlyLysArgThrLeuLys 774
 Db 2462 GCTAGTGTTTAAGAAAGGCAAGAGACCTCTAAG 2495

RESULT 4

BC028897 2656 bp mRNA linear HTC 04-AUG-2004
 LOCUS Mus musculus cDNA clone IMAGE:4168013, containing frame-shift errors.
 ACCESSION BC028897
 VERSION BC028897.1 GI:22137711
 KEYWORDS HTC
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

REFERENCE

1 (bases 1 to 2656)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B.S., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smalick, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 2656)
 Director MGC Project.
 Direct Submission
 Submitted (04-MAY-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center

Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Manavathi,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 24 Row: f Column: 2
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 31560679
 This clone has the following problem: frame shifted.

FEATURES

source

1..2656
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /db_xref="IMAGE:4168013"
 /tissue="types:salivary gland, 10 week old female mouse"
 /clone_lib="NCI CGAP_SG2"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:
 Pred. No.: 3,48e-238 Length: 2656
 Score: 2200.00 Matches: 417
 Percent Similarity: 73.82% Conservative: 113
 Best Local Similarity: 58.08% Mismatches: 141
 Query Match: 54.37% Indels: 47
 DB: 3 Gaps: 7

US-10-028-384-8 (1-774) x BC028897 (1-2656)

QY 17 SerLeuIleThrPheAlaIleLeuLeuLeuAlaIlePheSerArgLeu 36
 Db 65 ACATCTTAAAGCTTCTCATCTCTCGATGGCTGCTGTATCTTTTCTACTCGCTT 124
 QY 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProThrPheAsnThrArg 56
 Db 125 TTCTGCTGCTGAGATTGAAGATGTCATCCATGAGTTGATCCGACTTTTAATATCGG 184
 QY 57 AlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArg 76
 Db 185 ACTACCCGGTTCTCGCTGAGGAGGGGTTTATAATTCATTAATCGGTTGATGACCG 244
 QY 77 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
 Db 245 GCTTGGTACCTTTGGGCCGGAATCATGGAGGAAACAATTTACCCAGGTTTAATGATCCT 304
 QY 97 SerGlyGlyIleHisTrrLeuLeuHisValLeuAsnIleProValHisIleArgAspIle 116
 Db 305 TCTGCTGCATCTACCATGATGATCCATCTTCTCCATATCATATTGACATTCGGAATGTC 364
 QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
 Db 365 TGTGTTTCTCGGCCCACTTTTCTCTCTTTCACCCACCATCGTTACGACCTTACC 424
 QY 137 LysGluLeuTrrSerAlaGlyValGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
 Db 425 AAAGAGCTCAAGGATGAGAGGCTGGGCTTCTGCTGCTGCCATGATTGCTGTAGTTCT 484
 QY 157 GlyTyrIleSerArgSerValAlaGlySerTyAspAsnGluGlyIleAlaIlePheAla 176
 Db 485 GGGTATATTTCTCATGCTAGTGGCTCTCTATGATAATGAAGAAATTCGTATCTTTTC 544
 QY 177 LeuGlnPheThrTyrPheLeuTrrValArgSerValIysThrGlySerValPheTrpSer 196
 Db 545 ATGCTGCTTACTTACTACATGATGATCAAGGAGTGAAGACTGTTCCATCTATTGGGCT 604
 QY 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 216

605 GCCAAGTGGCCCTCGCTTATTTCTACATGGTCTCTTCATGGGAGGCTAATGTTCTCGT 664
 217 ILeAsnLeuLeuProLeuHisValPheValLeuLeuLeuMetGlyArgTySerProArg 236
 665 ATCAACTGATTCCTCTACATGCTCTGGTGGCTAAGCTGACAGGCGGTTTTCACCGG 724
 237 LeuLeuThrSerTySerThrPheTyriLeuLeuGlyLeuPheSerMetGlnIlePro 256
 725 ATCTACGTAGCTACTGTACTGTTTACTGCTGGGCAACATCTTCTATGACAGATTCC 784
 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
 785 TTTGTTGGTTTCAGGCGGCTCTTTCATCAGACACATGCGGCTTTCGGAGTGTGGT 844
 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
 845 CTCTGTGAGATCCATGCTTTCGTAGATTACCTGCGGAGCAAGTTGAATCCACAGCAATC 904
 297 ArgLeuLeuPhe 911
 905 GAAGTCTTTTCGGAGTGTATCTCCCTGGTGGCTTGTCTCTCTCACTGTGGGA 961
 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProTyrSerGlyArgPhe 331
 962 -----GCTCTCTCATGCTAAC-----GGAAAAATTTCTCCCTGGACAGGCGGTTTC 1009
 332 TySerLeuTrpAspThrGlyTyriAlaLysIleHisIleProIleIleAlaSerValSer 351
 1010 TACTCTCTGTCGATCCCTCTTATGCTAAGATAACATTCCTCATTTATGCTATCTGTTCT 1069
 352 GluHisGlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla 371
 1070 GAGCACCAGCCCAACCTGCTCTCTACTATTTTGTATCTACACCTCTCTGTTCTCATG 1129
 372 PheProValGlyValTrpTyriCysIleLysGlnIleAsnAspGluArgValPheValVal 391
 1130 TTTCCAGTTGGCTCTTACTGCTTTAGCAACCTGCTGTGATGCTGCGGATTTTATCATC 1189
 392 LeuTyriAlaIleSerAlaValTyriPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
 1190 ATGATGCTGTGACGAGCATGCTATTTTCAGCTGTATGTTGGTGGCTCTATGCTGTTATG 1249
 412 ThrProValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
 1250 GCACCTGTATGTCATCTTCTGTCATGTTGTTCCAGTCTG----- 1297
 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
 1297 ----- 1297
 452 AlaGluAspSerIleGluLysLysThrLeuTyriAspLysAlaGlyLysLeuLysHisArg 471
 1298 -----TCCATATGAAAAATCTGGACATAATGCCCCAGA 1333
 472 -ThrLysHisAspAlaGlnAspThr-----GlyValSerAsnLeuLysSerIle 489
 1334 CAAGAAGAGCAAGCAAGCAAGCATCTTACCTATTAAAGATGAGGTGGCGAGTGG 1393
 489 eValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValTh 509
 1394 GATGATCTGCTGCTGCTTTTCTCATCCTACCTACACGTTTCAATCGATCGGTGGTGC 1453
 509 rSerAsnAlaTyriSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
 1454 CAGTGAAGCCTATTCTCTCCCTCATCTTGTCTGCTGCTGCTGGG-----GATGCGAG 1510
 529 rArgAsnIleLeuAspAspPheArgGluAlaTyriTyriProLeuSerGlnAsnThrAlaAs 549
 1511 TAGATCAITTTTGTGATCTTCGAGAGCGTATTATGCTGCTCCGTCATCATCTCCAGA 1570
 549 PaspAlaArgValMetSerTrpAspTyriGlyTyriGlnIleAlaGlyMetAlaAsnAr 569

1571 GGATGCAAAAGTCATGTCATGTTGGGATTATGCTGCTACCAATTTACTGCAATGGCAATCG 1630
 569 gThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAl 589
 1631 GACAAATTTTAGTGGCAATAACACATGGAATAATACCATATTTCTCGATGGGAGGC 1690
 589 aMetSerSerThrGluGluLysSerTyriGluIleMetThrSerLeuAspValAspTyriVa 609
 1691 AATGCGATCCACAGAAAGAAAGCTATGAATCATGAGGAGCTTGATCTCAGCTATGT 1750
 609 ILeuValIlePheGlyGlyValIleGlyTyriSerGlyAspIleAsnLysPheLeuTr 629
 1751 GCTTGTGATTTTGGAGCCCTTACTGGGTATTTCTCGATGATATCAACAAGTTCTTTG 1810
 629 pMetValArgIle---AlaGluGlyGluHisProLysAspIleLysGluSerAspTyriPh 648
 1811 GATGTCGGATTGGAGGAAGACAGAGACAGAGACAGGAAGACACATTAAAGAGAATGACTACTA 1870
 648 eThrAspArgGlyGluPheArgValAspAlaGluGlyValAlaProAlaLeuLeuAsnCysLe 668
 1871 TACTCTCTGCTGGGAATTCCTGTTGATCGTGGGTTCTCCGGTCTGCTCAACTGCCT 1930
 668 uMetTyriLysLeuSerTyriTyriArgPheGlyGluLeuLysLeuAspTyriArgGlyProSe 688
 1931 TATGTACAAATGTTGTTACTACCTGTTTGGCAGGTCTACACAGAGCCAGCGTCCACC 1990
 688 rGlyTyriAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyriLe 708
 1991 AGCTTTTACCGCTGTGCAATGCTGAGATTGGTAATAAAGACTTTGAGGCTGATGCTCT 2050
 708 uGluGluAlaTyriThrThrGluHisTrpLeuValArgIleTyriArgValLys 725
 2051 GGAGGAGCGTATATACACAGAACACTGGTAGTCAGGATATACAGGTAAG 2102

RESULT 5

AK030363 3802 bp mRNA linear HTC 03-APR-2004
 LOCUS Mus musculus adult male pituitary gland cDNA, RIKEN full-length
 DEFINITION enriched library, clone:530402C17 product:OLIGOSACCHARYL
 TRANSFERASE STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
 homology (Mus musculus), full insert sequence.
 ACCESSION AK030363.1 GI:26326354
 VERSION AK030363.1
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1
 REFERENCE Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 2
 JOURNAL 20499374
 MEDLINE 11042159
 PUBMED 11042159
 3
 REFERENCE Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitesunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Matsura, S., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multichannel sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagawa, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Submitted (16-Apr-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/

FEATURES

source

1. .2730
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:C130038C18"
/db_xref="taxon:10090"
/clone="C130038C18"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
125..2243

misc_feature

/note="OLIGOSACCHARYL TRANSFERASE STU3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1), homolog [Mus musculus] (SWISSPROT) P46978, evidence: FASTA, 99.%ID, 100%length, match=2115)
2717..2722
putative"

polyA_signal

2717..2722

polyA_site

2730

ORIGIN

Alignment Scores:

Pred. No.: 1,626-236 Length: 2730
Score: 2185.50 Matches: 416
Percent Similarity: 73.68% Conservative: 113
Best Local Similarity: 57.94% Mismatches: 143
Query Match: 54.02% Indels: 46
DB: 3 Gaps: 8

US-10-028-384-8 (1-774) x AK081547 (1-2730)

QY 17 SerLeuIleThrPheAlaIleLeuLeuLeuLeuAlaIlePheSerSerArgLeu 36
DB 170 ACACCTCTAAGCTTCTCATCTCTGTCGATGGCTGCTGGTATCTTTCTACTCGCTT 229
QY 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArg 56
DB 230 TTGCTGCTGCTGAGATTGGAAGTGCATCCATGAGTTGATCCGTACTTTAATTTATCGG 259
QY 57 AlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArg 76
DB 290 ACTACCCGGTTTCGCTCAGGAGGGGTTTATAATTCATACTGGTTTGATGACCGG 349
QY 77 AlaTTPTrpProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
DB 350 GCTTGATACCTTTGGCGCGAATCATTTGAGGAACAATTTACCCAGGTTTAATGATCACT 409
QY 97 SerGlyGlyIleHisTrpLeuHisValLeuAsnIleProValHisIleArgAspIle 116
DB 410 TCTGTCGCAATCTACCATGCTACTCCATCTTCTCCATATCACCATTGCAATTCGAATGTC 469

QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
DB 470 TGTGTTTCTTGGCCCGCTTCTCTCTTACCACCATGTTAGTACACCTTACC 529
QY 137 LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
DB 530 AAAGAGCTCAAGGATCAGGAGCTGGGCTTCTGCTGCTGCCATGATTGCTAGTTCT 589
QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
DB 590 GGGTATATTTCTGATCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 649
QY 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
DB 650 ATGCTGCTTACTTACTACATGTCATCAAGCAGTGAAGACTGGTTCATCTATGGGCT 709
QY 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 216
DB 710 GCCAAGTGTGCCCTGCTTATTTCTACATGCTCTTCAATGGGAGGCTATGTTCTGTG 769
QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlyArgTyrSerProArg 236
DB 770 ATCAACTTGATTCCTTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 829
QY 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
DB 830 ATCTAGGTAGCTTACTGTACTGTTTACTGCTGGGACCATTTCTTCTATGACAGATTCC 889
QY 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaLeuLeuGlyValPheVal 276
DB 890 TTGTTGTTTCCAGCCGCTCTTTCATCAGAACACATGCGAGCCTTTGGAGTGTGGT 949
QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
DB 950 CTCTCTCAGATCCAGCTTTCGTAGATCTCTGCGAGCAAGTGAATCCACAGCAATTC 1009
QY 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
DB 1010 GAAGTTCTTTTCCGAGGTGTTATCTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1066
QY 312 PheValAlaValValValLeuThrMetLeuGlyValValAlaProTrpSer-GlyArgPhe 331
DB 1067 -----GCTCTCTCTATGCTAACA-----GGAAAAATTTCTCCCTGGACAGGGGGGTTT 1114
QY 331 eTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSe 351
DB 1115 CTACTCTCTGCTGGATCCCTCTTATGCTAAGAAATACATTCCTCCATTTATGCTGTTTC 1174
QY 351 rGluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuValCysAl 371
DB 1175 TGACCCAGCCGCCACACCTGCTTCTACTATTTTGTACTAGCTCTCTGCTCTCTCTCAT 1234
QY 371 aPheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVa 391
DB 1235 GTTTCAGTTGGCTCTATTACTGCTTAGCAACCTGCTGATGCTCGGATTTTATCAT 1294
QY 391 lleuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLe 411
DB 1295 CATGATGCTGTGACCATGCTACTTTTACGCTGTATGCTGGTCTTAATGCTGTTAT 1354
QY 411 uThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLe 431
DB 1355 GGCACCTGTTATGTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1414
QY 431 uGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspG 451
DB 1415 GAATAATCTGCACATAGTCCCGCA-----1439
QY 451 uAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisAr 471
DB 1440 -----GACAAGAGAGCAAG-----1454

/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
(B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
(SWISSPROT)P46978, evidence: FASTV, 99.7%ID, 100%length,
match=2115)
putative"

ORIGIN

Alignment Scores:

Pred. No.: 1,32e-234 Length: 3643
Score: 2170.50 Matches: 415
Percent Similarity: 73.64% Conservative: 113
Best Local Similarity: 57.86% Mismatches: 144
Query Match: 53.65% Indels: 46
DB: 3 Gaps: 8

US-10-028-384-8 (1-774) x AK077877 (1-3643)

```

QY 17 SerLeuIleThrPheAlaIleLeuLeuIleAlaIleLeuAlaGlyPheSerSerArgLeu 36
DB 138 ACACCTCTAAAGCTTCTCATCTCTGTCGATGGCTGCTGTTATCTTTTCTACTCGTCT 197
QY 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArg 56
DB 198 TTTGCTGTCTGAGATTTTGAAGTGTCTATCCATGAGTTTGATCCGTACTTTAATTATCGG 257
QY 57 AlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPheLeuAsnTTPheAspGluArg 76
DB 258 ACTACCGGTTTCTGGCTGAGGAGGGGTTTATAAATTCATTAATCCATCTGGTTGATGACCG 317
QY 77 AlaTTPTrpProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
DB 318 GCTTGCTACCTTTTGGCGGATCATATGGAGAACAAATTTACCCAGGTTTAAATGATCACT 377
QY 97 SerGlyGlyIleHisTrpLeuHisValLeuAsnIleProValHisIleArgAspIle 116
DB 378 TCTGCTGCAATCTACCATGATCTCCATTTCTTCCATATCACCATTGACATCGGAATGTC 437
QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
DB 438 TGTGTTTCTGCTGGCCCGCTTTTCTCTCTTTCACCCACCATCGTTAGTACACCTTACC 497
QY 137 LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaLeuSerPheIleAlaIleValPro 156
DB 498 AAGAGCTCAAGATGCAGAGCTGGGCTTCTGCTGTCGATGATGCTGTAGTTCCT 557
QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
DB 558 GGGTATATTCTCGATCTGTAGCTGGCTCCTATGATAATGAAGAAATTTGCTATCTTTTGC 617
QY 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSer 196
DB 618 ATGCTGCTTACTTACTACATGTGGATCAAGGAGTGAAGACTGGTTCCATCTATGGGCT 677
QY 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaIleTrpGlyGlyTyrValPheIle 216
DB 678 GCCAAGTGTGCTCTGCTTATTCTACATGTGCTCTCTTCTGAGGAGGCTATGTGTTCTCT 737
QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg 236
DB 738 ATCAACTTGAATTCCTTACATGCTCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 797
QY 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
DB 798 ATCTAGCTAGCTTACTGTACTGTTTACTGCTGCTGGGACCAATTTCTTATGCAGATTTC 857
QY 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
DB 858 TTTGTTGGTTTCCAGCCCGCTCTTTCATCAGAACACATGGGAGGCTTTGGAGTGTTCGT 917
QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
DB 918 CTCTGTGATCCATCCATCTTCTAGATTACCTGCGCAGCAAGTTGAATCCACAGCAATTC 977

```

```

QY 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
DB 978 GAAGTTCTTTTCCGAGGTATTATCTCCTCGTGGTGGCTTTCTCTCTCACTGTGGGA--- 1034
QY 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProThrPheSerGlyArgPhe 331
DB 1035 -----GCTCTCTCTCATGCTAAACA-----GGAAAAATTTCTCTCGACAGCGCGTTTC 1082
QY 332 TyrSerLeuTrpAspThrGlyTyrAlaIleHisIleProIleAlaSerValSer 351
DB 1083 TACTCTCTCTGGATCCCTCTTATGCTAAGAATAACATCCCATTTATGCTATCTGTTCCT 1142
QY 352 GluHisGlnProThrThrTrpPheSerPhePhePheLeuHisIleLeuValCysAla 371
DB 1143 GAGCACAGCCCAACACCTGCTCTCTATTTTCTATTTTCTACAGCTCTTCTTCTCATG 1202
QY 372 PheProValGlyValTrpTyrCysIleGlnIleAsnAspGluArgValPheValVal 391
DB 1203 TTTCCAGTTGGCTCTATTACTGCTTTAGCAACCTGCTCATGCTCGGATTTTATCATC 1262
QY 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
DB 1263 ATGTATGTGTGACGAGCATGCTTTTCTAGCTGTAAATGTGGTCTTAATGCTGTATTG 1322
QY 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
DB 1323 GCACCTGTATTGTGATCTTTCTGGCATTTGCTTCCAGGCTGCTGCCATATATG 1382
QY 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
DB 1383 AAAAATCTGGACATAAGTCGCCCA----- 1406
QY 452 AlaGluAspSerIleGluLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
DB 1407 -----GACAAAGAGAGCAAG----- 1421
QY 472 ThrLysHisAspAlaGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 499
DB 1422 -----AGCAACAGGATCTACTTACCTATTAAAGATGAGGTGGGAGTGGG 1469
QY 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
DB 1470 ATGATACTGGTCATGGCTTTCTTCTCATCACCTACACGTTTCAATCCAGCTTGGTGACC 1529
QY 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
DB 1530 AGTGAAGCTATTCTTCTCCTCCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1586
QY 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAsp 549
DB 1587 AGGATCAATTTTGATGACTTCCGAGAGCGTATTATTGGCTCCGCTCAATATCTCCAGAG 1646
QY 550 AspAlaArgValMetSerTrpTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
DB 1647 GATGCAAAAGTCATGTCATGTGGGATTTGCTTACCAATTTACTGCAATGGCAATCGG 1706
QY 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
DB 1707 AA-ATTTAGTGGACAATAACACATGAATAATAATACCCTATTTCTCGAGTAGGCGCA 1765
QY 590 MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
DB 1766 ATGGCATCCACAGAGAAAGAGCCATGAATCATGAGGAGCTGTGATGTCAGCTATGTG 1825
QY 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrp 629
DB 1826 CTGTGCAATTTTGGAGGCTTACTGGGTATTCTTCCGATGATATCAACAAGTTTCTTTGG 1885
QY 630 MetValArgIle---AlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
DB 1886 ATGGTCCCGGATTTGGAGGAGGACAGACAGGAGACACATTAAAGGAGATGACTACTAT 1945
QY 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668

```

```

Db 1946 ACTCCTACTGGGAATTCCTGTTGATCGAGGGTCTCCGGTGTGCTCAACTGCCTT 2005
Qy MetTyrlsLeuSerTyrtYrArgpHeGlyGluLeuLysLeuAspTyrtYrArgGlyProSer 688
Db 2006 ATGTACAAATGTTACTACCTGTTGGCAGGCTCTACACAGAACCAAGCGTCCACCA 2065
Qy GlyTyrtYrAspArgTrArgAsnAlaValLleGlyAsnLysAspPheAspLeuThrtYrLeu 708
Db 2066 GGCCTTACCGGTGTTCGAATGCTGAGATTGGTAATAAAGACTTTGAGCTTGATGCTG 2125
Qy GluGluAlaTyrtYrThrtYrGluHisTrpLeuValArgLleTyrtYrArgValLys 725
Db 2126 GAGGAAGCGTATACACAGAACACTGGCTAGTACAGATATACAGGTAAG 2176

RESULT 8
AY418284 2118 bp DNA linear GSS 17-DEC-2003
LOCUS Homo sapiens ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY418284
VERSION AY418284.1 GI:39774244
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2118)
Clark.A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sinsky, J.J.,
Adams, M.D. and Cargill, M.
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
1..2118
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2118
/gene="ITM1"
/locus_tag="HCM6506"

ORIGIN
Alignment Scores:
Pred. No.: 1,01e-223 Length: 2118
Score: 2071.50 Matches: 393
Percent Similarity: 72.12% Conservative: 101
Best Local Similarity: 57.37% Mismatches: 140
Query Match: 51.20% Indels: 51
DB: 9 Gaps: 8

US-10-028-384-8 (1-774) x AY418284 (1-2118)
Qy 52 TrpPheAsnTyrtYrArgAlaThrtYrMetValGlnAsnGlyTrpTyrtYrAsnPheLeuAsn 71
Db 151 TACTTTAATATCGGACTACAGGTTCTCGCTGAGGAGGGGTTTATATAATTCATAC 210
Qy 72 TrpPheAspGluArgAlaTyrtYrProLeuGlyArgLleValGlyGlyThrtYrValTyrtYrPro 91
Db 211 TGGTTTATGACCGAGGCTGTGACCTTTTGGGACGAATCATTTGGAGGACAAATTTACCCCA 270

```

```

Qy 92 GlyLeuMetIleThrsGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProVal 111
Db 271 GGTTTAATGATCACTCTGCTCAATCTACCATGTGATCTCCATTTTTCACATCACCATC 330
Qy 112 HisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrsIleSer 131
Db 331 GACATTCCGAATGCTGTGTCTCTGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 390
Qy 132 ThrTyrtYrLeuThrsGlyLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPhe 151
Db 391 ACGTACCACTTTACCAAGAGCTCAAGAGTCCAGGGCTGGCTTCTTGTCTGCTGCCATG 450
Qy 152 IleAlaIleValProGlyTyrtYrIleSerArgSerValAlaGlySerTyrtYrAspAsnGluGly 171
Db 451 ATTGCTGTAGTTCCTGGATATATCTCCGATCTGGGCTGGCTCTCTATGATATGAAGAGG 510
Qy 172 IleAlaIlePheAlaLeuGlnPheThrtYrPheLeuTrpValArgSerValTyrtYrGly 191
Db 511 ATTGCCATCTTTTGCATGCTACTACCTACTACATGTGGATCAAGGCAGTAAAGACTGGT 570
Qy 192 SerValPheTrpSerAlaAlaAlaLeuSerTyrtYrPheTyrtYrMetValSerAlaTrpGly 211
Db 571 TCCATCTGTGGGACGCTAAGTGTGCCCTTGTATTTCTACATGGTCTCTGTCATGAGGA 630
Qy 212 GlyTyrtYrValPheIleAsnLeuIleProLeuHisValPheValLeuLeuMetGly 231
Db 631 GGTATGTCTCTGATCACTTAATCTCTCTCCAGCTCTCTGCTGCTGATGCTCACAGGC 690
Qy 232 ArgTyrtYrSerProArgLeuLeuThrsTrpTyrtYrSerThrtYrIleLeuGlyLeuPhe 251
Db 691 CGTTTCTCTCACCGGATCTATGTGCCCTACTGTTTACTGCTCTGCTGGGCACTATATCT 750
Qy 252 SerMetGlnIleProPheValGlyPheGlnProIleArgTrpSerGluHisMetAlaAla 271
Db 751 TCTATGCAGATCTCTTGTGGGTTTCCAGGCTGTCTCTTTCATCAGAGCACATGGCAGCC 810
Qy 272 LeuGlyValPheValLeuLeuMetAlaAlaThrtYrLeuArgHisLeuGlnSerValLeu 291
Db 811 TTTGGGCTCTTTGGTCTCTGCCAGATCCATGCTTGTGGATTTACCTGGCGGCAAGTTG 870
Qy 292 SerArgAsnGluPheArgGlyLeuPhe-----IleValGlyGlyLeuLeu 306
Db 871 AATCCACCAAAATTTGAAGTCTTTTCCGAGGCTCATCTCTCTGTTAGGTTTGTCTCT 930
Qy 307 ValGlyValGlyValPheValAlaValValLeuThrtYrMetLeuGlyValValAlaPro 326
Db 931 CTCACCGTGGGA-----GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 975
Qy 327 TrpSerGlyArgPheTyrtYrSerLeuTrpAspThrtYrAlaLysIleHisIleProIle 346
Db 976 TGGACGGGCGTTTCTACTCGTGTGATCCCTCTTATGCTAAGAACAAACATCCCCATC 1035
Qy 347 IleAlaSerValSerGluHisGlnProThrtYrPheSerPhePheAspLeuHis 366
Db 1036 ATTGCTCTGTCTGAGCATCAGCCCAACCTGGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1095
Qy 367 IleLeuValCysAlaPheProValGlyValTyrtYrCysIleLysGlnIleAsnAspGlu 386
Db 1096 CTCCTCGTCTCATGTTTCCAGTTCGCTCTATTTACTGCTTTAGCAACCTCTCTGTATGCC 1155
Qy 387 ArgValPheValValLeuTyrtYrAlaIleSerAlaValTyrtYrPheAlaGlyValMetValArg 406
Db 1156 CGGATTTTATCATGATGTTGACCGAGCATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1215
Qy 407 LeuMetLeuThrtYrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeu 426
Db 1216 CTAATCTAGTGTGGCACTGTTATGTGCACTTCTCTGCGCATTTGGAGTCTCCAGGTG 1275
Qy 427 LeuAspValPheLeuGlnLysAspSerSerLysArgMetGlyThrtYrAlaIleSerAlaAla 446
Db 1276 CTGTCACATACATGAAGAATCTGGACATAAGTCTGTCGA----- 1314

```



```

/db_xref="PANTOM_DB.E130302E03"
/db_xref="caxon:10090"
/clone="E130302E03"
/tissue_type="eyeball"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
108 - 2233
/misc_feature
/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
(BS) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
(SWISSPROT|P46978, evidence: FASTA, 99.%id, 100%length,
match=2115)
putative"
2708 - 2713
/polya_signal
/note="putative"
2727
/polya_site
/note="putative"
ORIGIN
Alignment Scores:
Pred. No.: 4,95e-223 Length: 2727
Score: 2067.00 Matches: 406
Percent Similarity: 72.15% Conservative: 112
Best Local Similarity: 55.79% Mismatches: 148
Query Match: 51.09% Indels: 63
DB: 3 Gaps: 9
US-10-028-384-8 (1-774) x AK087470 (1-2727)
QY 17 SerLeuIleThrPheAlaIleLeuLeuLeuAlaIleLeuAlaGlyPheSerSerArgLeu 36
DB 153 ACACITTTCTAAAGCTTCTATCTGCTGCGATGGTGTCTGTTATCTTTCTACTGCTT 212
QY 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArg 56
DB 213 TTGCTGTGCTGAGATTGAAAGTGTCACTCCATGATGTTGATCGTACTTAAATTATCGG 272
QY 57 AlaThrAlaTrpMetValGlnAsnGlyTrpTrpAsnPheLeuAsnTrpPheAspGluArg 76
DB 273 ACTACCCGGTTTCTGCTGAGGAGGGGTTTATAAATTCATTAATCCATTCGATTCGACCGG 332
QY 77 AlaTrpTrpProLeuGlyArgIleValGlyGlyThrValTrpProGlyLeuMetIleThr 96
DB 333 GCTTGTGTACCCITTTGGCCGAATGATGGAGGACAAITTAATCCAGGTTTAAATGATCACT 392
QY 97 SerGlyGlyIleHisTrpLeuLeuHisValLeuLeuAsnIleProValHisIleArgAspIle 116
DB 393 TCTGCTGCAATCTACCATCTACTTCCATTTCTTCCATATCACCATTCGATTCGGAATGTC 452
QY 117 CysValPheLeuAlaProIlePheSerGlyLeuTrpSerIleSerThrTrpLeuLeuThr 136
DB 453 TGTGTTTCTGCGCCCGCTTTCTCTCTTTCCACCACCATCTGTTACGTACACCTTACC 512
QY 137 LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
DB 513 AAAGAGCTCAAGATGAGAGCTGGGGTTCTTGTGTGGTGCATGATTCCTAGTTCCT 572
QY 157 GlyTrpIleSerArgSerValAlaGlySerTrpAspAsnGluGlyIleAlaIlePheAla 176
DB 573 GGGGATATTCTCGATCTGATCTGCTCTCTATGATATGAGGAATGCTATCTTTTTCG 632
QY 177 LeuGlnPheThrTrpPheLeuTrpValArgSerValIleThrGlySerValPheTrpSer 196
DB 633 ATGCTGCTTACTTACTACATGTGGATCAAGGAGTGAAGACTGGTTTCCATCTATTGGGCT 692
QY 197 AlaAlaAlaAlaLeuSerTrpPheTrpMetValSerAlaTrpGlyGlyTrpValPheIle 216
DB 693 GCCAAGGTGCTCGCTTATTTCTATCATGCTCTCTTCATGGGAGGCTATGTTTCCG 752
QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTrpSerProArg 236
DB 753 ATCACTTGTATCTCTACATGTCCTGCTGCTAATGCTGACAGCCGCTTTTCTACCCGG 812
QY 237 LeuLeuThrSerTrpSerTrpPheTrpIleLeuGlyLeuLeuPheSerMetGlnIlePro 256

```


QY 52 TrpPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsn 71
DB 151 TACTTTAATATTCGAGCTACCGGGTTCTGGCTGAGGAGGGTTTATATATTCCTAATAC 210
QY 72 TrpPheAsnGluArgAlaTyrTyrProLeuGlyArgIleValGlyThrValTyrPro 91
DB 211 TGGTTTGAAGCCGGCTGGTACCCCTTTGGCCGGAATCATTTGGAGGAACAATTTACCCA 270
QY 92 GlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProVal 111
DB 271 GGTTTAATGATCATTCTGCTGCAATCTACCATGCTACTCTCATTTCTCCATATCATCTATT 330
QY 112 HistLeuArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSer 131
DB 331 GACATTCGGAATGCTGGTATTTCTGCCCCACCTTTCTCTCTTTCACACCACCATCGTT 390
QY 132 ThrTyrLeuLeuThrLysGluLeuTrpSerAlaGlyValGlyLeuPheAlaAlaSerPhe 151
DB 391 ACGTACCACTTACCAAGAGCTCAAGATGCAGAGCTGGGCTTCTGCTGCTGCTGCTGCTG 450
QY 152 IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGly 171
DB 451 ATTCTGTAGTTCCTGGGTATTTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTG 510
QY 172 IleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGly 191
DB 511 ATTCTGTATCTTTCATGCTGCTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 570
QY 192 SerValPheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGly 211
DB 571 TCCATCATTTGGGCTGCAAGTGGCCCTCGCTTATTTCTTACTGCTCTCTTCTCATGGGA 630
QY 212 GlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGly 231
DB 631 GGCTATGTTCTTGGTATCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 690
QY 232 ArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPhe 251
DB 691 CGTTTTCTCACCGATCTACGTAGCTACTGATCTTACTGCTTACTGCTGCTGCTGCTGCT 750
QY 252 SerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAla 271
DB 751 TCTATGAGATTTCTTGTGTGTTCAGCCGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 810
QY 272 LeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeu 291
DB 811 TTTGGAGTGTGCTCTCTCTCAGATCCATGCTTCTGATATTTCTGCTGCTGCTGCTGCT 870
QY 292 SerArgAsnGluPheArgLysLeuPheIleValGlyLeuLeuValGlyValGlyVal 311
DB 871 AATCCACAGCAATTCGAAGTCTTTTCCGGAGTGTATCTCTCTGCTGCTGCTGCTGCT 921
QY 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
DB 921 TTTGCTCTCTCTCTGAGGAGCTCTCTCTCGA-AAAAATTTCTCTCTCTCTCTCTCTCTCT 980
QY 332 TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
DB 981 TACTCTCTGCTGATCCCTCTTATGTAAGATAACATTCCTATTCATATTCATCTGTTCT 1040
QY 352 GluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla 371
DB 1041 GAGCACAGCCCAACCTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1100
QY 372 PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
DB 1101 TTTCCAGTTGGCTCTATTACTGCTTTAGCAACCTGCTGATGCTCGGATTTTATCATC 1160
QY 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValValLeuMetLeuThrLeu 411
DB 1161 ATGTATGGTGTACCAAGCATGACTTTTTCAGCTGTAATGGTGGCTGCTAATGCTGTTATG 1220
QY 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431

DB 1221 GCACCTGTTATGGCAATCTTTCTTGGCAATGGTGTTCCTCCAGTGTCTGCCACATATATG 1280
QY 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
DB 1281 AAAAATCTGACATAAGTCGCCCA----- 1304
QY 452 AlaGluAspSerIleGluLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
DB 1305 -----GACAAGAGAGCAAG----- 1319
QY 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
DB 1320 -----AAGCAACAGATTTCTACTTACCTATTAAGAAATCAGGTGGCGAGTGGG 1367
QY 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
DB 1368 ATGATACTGGTCAATGGCTTTTCTCTCCCTCCATTTGCTACTGCTGCTGCTGCTGCTGCTG 529
QY 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 549
DB 1428 AGTGAAGCCTATTTCTCTCCCTCCATTTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTG 1484
QY 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAsp 549
DB 1485 AGGATCATTTTGTATGACTTCCGAGAAGCGTATTATTGGCTCCGTCACAACTATCCAGAG 1544
QY 550 AspAlaAtgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
DB 1545 GATGCAAGAAGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1604
QY 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
DB 1605 ACAATTTTAGTGACAATAACACATGGAATAATATCCCATATTTCTCGAGTAGGCGAGCA 1664
QY 590 MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
DB 1665 ATGGCATCCACAGAAAGAAAGCCATGAAATCATGAGGAGCTGTGATGTGCTGCTGCTGCTG 1724
QY 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 629
DB 1725 CTTGTCAATTTTGGAGGCTTACTGGGTATTTCTTCGGATGATATCAACAAGTTTCTTTGG 1784
QY 630 MetValArgIle---AlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
DB 1785 ATGTCTCCGATTTGGAGGAGCAGACAGACAGACACATTAAGAGGAATGACTACTAT 1844
QY 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
DB 1845 ACTCTACTCTGGGAAATTCCTGTTGATCGTGAGGGTCTCCGCTGCTGCTCACTGCTCTT 1904
QY 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLysLeuAspTyrArgGly-Prose 688
DB 1905 ATGTACAAAATGTCTTACTACCGCTTTGGGAGCTCTACACAGAA-----GCCTCCACC 1958
QY 688 rGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLe 708
DB 1959 AGGCTTTGACCGTGTTCGAATGCTGAGATTGGTAATAAAGACCTTTCAGCTTGTATGCTCT 2018
QY 708 uGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLys 725
DB 2019 GGAGGAAGCGTATACACAGACACACTGGCTAGTCAGGTATACAGGTAAAG 2070

RESULT 12

CNS0ADYZ

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

CNS0ADYZ 2419 bp mRNA linear HTC 06-FEB-2004
Arabidopsis thaliana Full-length cDNA Complete sequence from clone
GSLTPEG51ZC09 of Hormone Treated Callus of strain col-0 of
Arabidopsis thaliana (thale cress).

BX816490

BX816490.1

GI:42474670

HIC; GSLT_cDNA.

Arabidopsis thaliana (thale cress)


```

QY 442 AlalleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThrLeu 461
Db 1452 ACTATCAAGATTAACTCTCTGTTAAGGACAAACAAAGGTTTCTCACACTGGTTC 1511
QY 462 TyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThrGly 481
Db 1512 ACGAAAGGAGCTGGTAGTTCAAAGCTTCTTCAAGGTTTACACTTGATCAGTCTCAGCCT 1571
QY 482 ValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuMetMetPhe 501
Db 1572 TTCCAGAGAAATGGTGCCATGCTCTCTGTTGTTGTTATTTTTCCTCAGTAGATAT 1631
QY 502 AlavalHisCysThrTyrValThrSerAsnAlaTyrSerProSerIleValLeuAla 521
Db 1632 GCTATTCTACGTCACATGGGTGACAGCAGGACATATTCTCTCCTCAATTGCTTAGCT 1691
QY 522 PheHisAsnSerGlnAspGlySerArgAsnIleLeuAspPheArgGluAlaTyrTyr 541
Db 1692 GCNAGAGAGGCCAC--GGGAACAGAAATCATCTTTGATGATTACCGGAGGCATATC 1748
QY 542 TrpLeuSerGlnAsnThrAlaAspAlaArgValMetSerTrpTyrAspTyrGlyTyr 561
Db 1749 TGGCTTAGGCAAAACACATGCCACTGATGCTAAGATTATGCTGTTGGGACTATGGTAC 1808
QY 562 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 581
Db 1809 CAATCTACCTGCGTGGGAATAGAACTGTCTATCGTATGATAACAATACCTGGAAACAAC 1868
QY 582 HisIleAlaLeuValGlyLysAlaMetSerThrGluGluLysSerTyrGluIleMet 601
Db 1869 CACATTGCTACCGTTGGAGCCGCCATGCTCTTCTTATGAGATGATGCTGATGATCATG 1928
QY 602 ThrSerLeuAspValAspTyrValLeuValIlePheGlyValIleGlyTyrSerGly 621
Db 1929 AGGTCATCTGATGTAAGTATGTTGCTGCTCTTTGCTGTTGTTGTTGTTGTTGTTTCA 1988
QY 622 AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp 641
Db 1989 GATGATATCAACAAGTCTTGTGATGTTGGAGATTGGAGTGGAGTATTCACAGTG-- 2045
QY 642 IleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAla 661
Db 2046 ATCAAGAACCTGATTAACCTTGTGAAT--GGTGAATTCCTGTAGACAAAGTGATCA 2102
QY 662 ProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLys 681
Db 2103 CCGAAGATGCTGAACCTGCTCATGTACAAAGTTATGCTATTACAGGTTCGGTGAAGTACC 2162
QY 682 LeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLys 701
Db 2163 ACAGATACGCAAGCCACCCGGTATGATAGACGAAGAGAGTGGAGATAGGGAACAA 2222
QY 702 AspPheAspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIle 721
Db 2223 GATATCAAACTTGAACTTAGAAGAGCTTACAAACATCCAACTGATGATGTTGTTAT 2282
QY 722 TyrArgValLysLysPro 727
Db 2283 TACAGAGTTAAACCTCT 2300

RESULT 13
LOCUS AY418285 2079 bp DNA linear GSS 17-DEC-2003
DEFINITION Pan troglodytes ITM1 gene, VIRUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY418285
VERSION AY418285.1 GI:39774245
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 2079)

```

AUTHORS

Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J., Adams, M.D. and Cargill, M.
Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios

JOURNAL

Science 302 (5652), 1960-1963 (2003)

PUBMED

14671302

REFERENCE

2 (bases 1 to 2079)

AUTHORS

Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J., Adams, M.D. and Cargill, M.
Direct Submission

JOURNAL

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
These sequences were made by sequencing genomic exons and ordering them based on alignment.

FEATURES

Location/Qualifiers
source
1..2079
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>2079
/gene="ITM1"
/locus_tag="HCM6506"

ORIGIN

US-10-028-384-8 (1-774) x AY418285 (1-2079)

Alignment Scores:

Pred. No.: 5,46e-178 Length: 2079
Score: 1669.50 Matches: 332
Percent Similarity: 60.57% Conservative: 75
Best Local Similarity: 49.40% Mismatches: 234
Query Match: 41.26% Indels: 31
DB: 4 Gaps: 4

QY 53 PheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrp 72
Db 154 TTTAATTATCGACTACCAAGNNCTGCTGNNNNNGGTTTATAAATTCATAACTGG 213
QY 73 PheAspGluArgAlaTyrTyrProLeuGlyArgIleValGlyGlyThrValTyrProGly 92
Db 214 TTGTGATCCGAGCTGCTGCTACCTTTGGAGCAATCATTTGGAGGAGCANNNTTACCCAGGT 273
QY 93 LeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHis 112
Db 274 TTAATGATCCTCTGCTGCAATCTACCATGCTACTTCATTTTCCACATCACCATGAC 333
QY 113 IleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThr 132
Db 334 ATTCGAATGCTTNTGTCTGCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 393
QY 133 TyrLeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIle 152
Db 394 TACCACCTTCAAGAAGCTCAAGATGCAAGGCTGGGCTTCTGTGCTGCTGCTGCTGCT 453
QY 153 AlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIle 172
Db 454 GCTGTAGTTCCTGGATATATCTCCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 513
QY 173 AlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySer 192
Db 514 GCCATCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 573
QY 193 ValPheTrpSerAlaAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGly 212
Db 574 ATCTGTTGGGAGCTAAGTGTGCCCTTGTCTTATTTCTACATGCTGCTGCTGCTGCTGCT 633
QY 213 TyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArg 232
Db 634 TATGTGTCTCTGATCAACTTAATTCCTCCACGCTCCTGCTGCTGCTGCTGCTGCTGCT 693

QY 233 TysrProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSer 252
 Db 694 TTCTCTACCGGATCATGTGGCCCTACTGTACTGTCTTACTCGCTGGGCACTACTTTCT 753
 QY 253 MetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeu 272
 Db 754 ATGCAGATCTCTTTGTGGGTTTCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 813
 QY 273 GlyValPheValLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSer 292
 Db 814 NNN 873
 QY 293 ArgAsnGluPheArgLysLeuPheIleValGlyLeuValGlyValGlyValPhe 312
 Db 874 CCANAACAATTTNNNGTTCTTTCCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 933
 QY 313 ValAlaValValLeuThrMetLeuGlyValValAlaProTyrSerGlyArgPheTyr 332
 Db 934 NNN 993
 QY 333 SerLeuTyrAspThrGlyTyrAlaLysIleHisIleProIleAlaSerValSerGlu 352
 Db 994 TCACCTCGTGGATCTTCTTATGTAAGAACAAACATCCCATCATTTGCTGTCTGCTGAG 1053
 QY 353 HisGlnProThrThrTyrPheSerPhePheAspLeuHisIleLeuValCysAlaPhe 372
 Db 1054 CATCACCCACACCTGGCTCTACTATTGTACCTNNAGCTCTCTCTCTCTCTCTCT 1113
 QY 373 ProValGlyValTyrPyrCysIleLysGlnIleAsnAspGluArgValPheValLeu 392
 Db 1114 CCAATTGGCTCTATTACTGCTTTNGCAACCTGTCTGATGCCGATTTTATCATCATG 1173
 QY 393 TyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThr 412
 Db 1174 TATGGTGTGACGACGATCTTTTTCAGCTGTAATG----- 1209
 QY 413 ProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeuGln 432
 Db 1210 -----NNNNNN 1215
 QY 433 GluAspSerSerLysArgMetGlyThrAlaIleSerAlaThrGluValAspGluAla 452
 Db 1216 NNN 1275
 QY 453 GluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArgThr 472
 Db 1276 NTGTCNNATACATGAAGAAANTGNACATAANNNNNNNNNNNNNNNNNNNNNNNN 1326
 QY 473 LysHisAspAlaGlnAspThrGlyValSerSerAsnLeuLysSerIleValleLeu 492
 Db 1327 AAGAAGNNNNAGGATCCCCNCCCTATTAAAGATGAAGTGGCAAGTGGGATGATCTG 1386
 QY 493 AlaValLeuMetLeuLeuMetPheAlaValHisCysThrTyrValThrSerAsnAla 512
 Db 1387 GTACGTGCTTTCTTCTCATCACCTACCTTCATTCACCTGGGTGACCGATGGGCC 1446
 QY 513 TyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIle 532
 Db 1447 TACTTCTTCGTCCTCATTTGATCTATCTGCGCGTGGTGGG---GATGGCAGTAGGATCAT 1503
 QY 533 LeuAspAspPheArgGluAlaTyrTyrTyrLeuSerGlnAsnThrAlaAspAlaArg 552
 Db 1504 TTTGATGACTCCGAGAGGATATTATTGCTCTGCTCACAATCTCCAGAGGATGGGAAG 1563
 QY 553 ValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeu 572
 Db 1564 GTCATNNCCITGGTGGATATTGGCTATCAGCTATGCGCAACCGAACAAATTTTA 1623
 QY 573 ValAspAsnAsnThrTyrAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 592
 Db 1624 GTGGACATACACATGATGAATAATACCCATATTCTCTNNNNNNNNNNNNNNNNNNNNNN 1683

QY 593 ThrGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIle 612
 Db 1684 ACAGAGAAAAGCCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTGCTATT 1743
 QY 613 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetValArg 632
 Db 1744 TTGGAGGCTCTCCTGGTATTCTCTGATGATATCAACAAGTTCTNNNNNNNNNNNNNN 1803
 QY 633 Ile---AAAGLuglyLysHisProLysAspIleLysGluSerAspTyrPheThrAspArg 651
 Db 1804 ATTGGAGGAGNACAGANNANGCAACATATCAAGGAGATGATATATATCTCACT 1863
 QY 652 GlyGluPheArgValAspAlaGluAlaProAlaLeuLeuAsnCysLeuMetTyrLys 671
 Db 1864 GGGAGTTCCGTTGNNCNCNTGAGTTCTCCAGTGTCTCANNNNNNNNNNNNNNNNNN 1923
 QY 672 LeuSerTyrTyrArgPheGlyLysLeuLysLeuAspTyrArgGlyProSerGlyTyrAsp 691
 Db 1924 NNN 1983
 QY 692 ArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeuGluAla 711
 Db 1984 CGTCTCGAAATGCTGAGATTGGGATAAAGACTTTGACCTTGATGCTCCTGGAGGAAGCA 2043
 QY 712 TyrThrThrGluHisTyrLeuValArgIleTyrArg 723
 Db 2044 TATACACAGAACATTTGGTGTGCTCAGGATATACAAG 2079

RESULT 14
 CNS017ZX 836 bp DNA linear GSS 26-JUL-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
 DEFINITION BAC37M18 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL108711.1 GI:5629015
 VERSION AL108711.1
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 836)
 REFERENCE Genoscope.
 AUTHORS Direct Submission
 TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (BDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES
 source
 1..836
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BAC37M18"
 /clone_lib="DrosBAC"
 /plasmid="pBelobAC11"
 /note="end : T7"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.66e-147 Length: 836
 Score: 1396.00 Matches: 274
 Percent Similarity: 98.92% Conservative: 0
 Best Local Similarity: 98.92% Mismatches: 2
 Query Match: 34.50% Indels: 1

us-10-028-384-8.rst

Thu Dec 16 16:25:17 2004

```

QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db 349 CATGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATTCGGTGTTCCTG 408
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
Db 409 GCGCGGANTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGTGG 468
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db 469 TCCGCGGGCGCGGCTCTTCGCGGCCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT 528
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db 529 AGTCGGTGGCTGGATCGTACGATACGAGGGCATTCGCATATTCGCCCTGCAGTTCACC 588
QY 181 TyrPheLeuTrpValArgSerValIlyThrGlySerValPheTrpSerAlaAlaAla 200
Db 589 TACTTCCTGCTGGTGGCTCAGTGAAGACTGGATCCGTGTTCTGGTGGCGCGCAGCGCT 648
QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle 220
Db 649 TTATCCTACTTCTACATGGTGCCGCTCCGCTGGGTGGCTAGTGTTCATCAACCTGATA 708
QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
Db 709 CCCCTGCACGCTCTTCGTACTGCTCATATTATGGCAGGTACTGCGCGCTGCTGTCACCAGC 768
QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGln 254
Db 769 TACAGCACCTTCTACATNCTGGGACTGCTGTTCTCCATGCAG 810

```

Search completed: December 15, 2004, 11:53:25
Job time : 6096.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 23:35:31 ; Search time 887.605 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 705

Sequence: 1 MTRKGFRLSYEKQDTLLKL.....HMLVRIYKVDLNLRLSRT 705

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8161859

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame, p2n.model DEV=xlh
-Q/cgn2_1/USFO_spc01/US10028384/runat_14122004_131648_6892/app_query.fasta_1.3740
-DB=published Applications NA -QFW=fastap -SUFFIX=Oligo.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10028384@cgn1 1.2124 @runat_14122004_131648_6892
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=10 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-Fgapop=6 -Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

Result No.	Score	Query Match	Length	DB ID	Description
1	705	100.0	2472	15	US-10-171-581-112
2	705	100.0	2472	15	US-10-028-384-112
3	705	100.0	2472	15	US-10-173-118-742
4	705	100.0	2472	16	US-10-343-887-742
5	226	32.1	3046	18	US-10-417-375-95
6	226	32.1	5404	18	US-10-417-375-99
7	226	32.1	5827	18	US-10-417-375-97
8	152	21.6	3093	18	US-10-417-375-92
9	152	21.6	3094	15	US-10-028-384-9
c 10	146	20.7	440	9	US-09-878-722-59
c 11	146	20.7	440	10	US-09-904-456-59
12	140	19.9	452	9	US-09-878-178-1456
13	140	19.9	452	13	US-10-046-935-1456
14	140	19.9	452	14	US-10-146-502-1456
15	128	18.2	487	10	US-09-871-161-213
16	126	17.9	575	9	US-09-525-297-53
17	102	14.5	487	9	US-09-736-457-1655
18	102	14.5	487	9	US-09-302-941-1655
19	102	14.5	487	9	US-09-849-626-1655
20	102	14.5	487	14	US-10-017-754-1655
21	102	14.5	487	15	US-10-113-872-1655
22	102	14.5	487	15	US-10-283-017-1655
23	101	14.3	307	9	US-09-736-457-375
24	101	14.3	307	9	US-09-736-457-1271
25	101	14.3	307	9	US-09-302-941-375
26	101	14.3	307	9	US-09-302-941-1271
27	101	14.3	307	9	US-09-849-626-375
28	101	14.3	307	9	US-09-849-626-1271
29	101	14.3	307	10	US-09-476-300-375
30	101	14.3	307	14	US-10-017-754-375
31	101	14.3	307	14	US-10-017-754-1271
32	101	14.3	307	15	US-10-113-872-375
33	101	14.3	307	15	US-10-113-872-1271
34	101	14.3	307	15	US-10-283-017-375
35	101	14.3	307	15	US-10-283-017-1271
36	94	13.3	419	9	US-09-360-352-10293
37	81	11.5	596	11	US-09-876-143-200
38	81	11.5	596	11	US-09-876-143-288
39	80	11.3	476	10	US-09-918-995-22164
40	65	9.2	307	9	US-09-736-457-588
41	65	9.2	307	9	US-09-302-941-588
42	65	9.2	307	9	US-09-849-626-588
43	65	9.2	307	10	US-09-476-300-588
44	65	9.2	307	14	US-10-017-754-588
45	65	9.2	307	15	US-10-113-872-588

ALIGNMENTS

RESULT 1
US-10-171-581-112
; Sequence 112, Application US/10171581
; Publication No. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10/171,581
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 112
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: L38961

1 DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-112

Alignment Scores:

Pred. No.: 0 Length: 2472
Score: 705.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-12 (1-705) x US-10-171-581-112 (1-2472)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 107 ATGACTAAGTTGGATTGTTGGGATTGCTCTATGAGAGCAGGACACACTTTTGAAGCTT 166
Qy 21 LeuileLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGCAATGGCTGCTGTATTATCTCTCCACTCGTCTGTTGCTGCTGCTGAGA 226
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTTGAAAGTGTATPCCATGAGTTGATCCGTACTTTTAAATTATCGGACTACCAAGTTCTCTG 286
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTrpTyrProLeu 80
Db 287 GCTCAGGAGGGGTTTATAAATCCATAACTGGTTTGATGACCGAGCTGGTACCCCTTGTG 346
Qy 81 GlyArgIleLeuGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGAGCAATCATTTGGAGGAACAATTTATCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 406
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTCCACATCACCATCGACATCGCAATGCTGTGTGTCTCTGCTGCC 466
Qy 121 ProLeuSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CCT 526
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
Db 527 GCAGGGCTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 586
Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGTGGCTGCTCTCTATGATTAAGAGGATTGCCCATCTTTGATGCTACTACCTAC 646
Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGATCAAGGCAGTAAAGATGGTTCCATCTGTGGCAGCTAAGTGTGCCCTT 706
Qy 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIleAsnLeuIlePro 220
Db 707 GCTTATTCTACARGTCTCTGATGGGAGGTTATGTTTCTGATCAACTTAAATCTCT 766
Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db 767 CTCACAGTCTCTGCTGATGCTCACAGGCGGTTTCTCTCACCGGATCTATGTGGCCTAC 826
Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 827 TGTACTGTTTACTGCTGGTACTATATCTTTCTAGCAGATCTCTCTGTGGGTTTCCAG 886
Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 887 CCTGCTCTTTCATCAGAGCATGGCAGGGGTTTGGGGTCTTTGGTCTCTGCCAGATCCAT 946
Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArg 300
Db 947 GCCTTTGTGGATTACTTGGCAGCAAGTTGAATCCACACNAATTTGAAGTCTTTTCCGG 1006

Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db 1007 AGCGTCATCTCTCTGGTAGGCTTTGCTCTCTCACCGTGGGAGCTCTCTCTCATGTCGACA 1066
Qy 321 GlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db 1067 GGAATAATATCTCTCTGACGCGGGCTTTCTACTCACTGCTGGATCCCTCTTATGTAAG 1126
Qy 341 AsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
Db 1127 AACCAATCCCCATCANTGCTCTGCTGAGCATCAGCCCAACCTGGTCTCTCATAC 1186
Qy 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
Db 1187 TATTTTGACCTGACGCT 1246
Qy 381 AsnLeuSerAspAlaArgIlePheIleLeuMetTyrGlyValThrSerMetTyrPheSer 400
Db 1247 AACCTGTCTGATGCGCGGATTTTATCATCATGATGATGCTGACCAAGCATGTACTTTCA 1306
Qy 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
Db 1307 GCTGTAATGGTGGCTCTAATGCTAGTGTGGACCTCTTATGAGCATTTCTCTGGCAIT 1366
Qy 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
Db 1367 GGAGTCTCTCCAGGTGCTGTCACATACATGAAGAATCTGGACATAAGTCCGCCAGACAAG 1426
Qy 441 LysSerLysLysGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
Db 1427 AGAGCAGAGACACACAGATTCACCTACCTATTAAGATTGAGTGGCAAGTGGGATG 1486
Qy 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
Db 1487 ATACTGGTCATGGCTTTCTTCTCATCACCTACACCTTTCATTCAACCTGGTGACCACT 1546
Qy 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
Db 1547 GAGGCTACT 1606
Qy 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluLysAla 520
Db 1607 ATATTGTATGATCTCCGAGAGCATATTATTTGGTCTCTCATATACTCCAGAGGATGGC 1666
Qy 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
Db 1667 AAGTTCATGCTCTGTGGGATTTATGGCTATCAGATTACAGTATGGCAACCGAACAT 1726
Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1727 TTAGTGGCAATAACACATGGAATAATACCATATTTCTGAGTAGGCGAGGCAATGGCG 1786
Qy 561 SerThrGluGluLysAlaTyrGluMetArgGluLeuAspValSerTyrValLeuVal 580
Db 1787 TCCACAGAGGAAAGCCTATGAGTATGAGGAGCTCGTGTGCTAGTATGCTGTGCTG 1846
Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
Db 1847 ATTTTGGAGGCTCTCACTGGGTATCTCTGATGATATCAACAGTTTCTTTGGATGGTC 1906
Qy 601 ArgIleGlyLysSerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
Db 1907 CGGATTTGGAGGAGGACAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCCA 1966
Qy 621 ThrGlyGluPheArgValAspGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1967 ACTGGGAGGATTCGGTGTGGACCGGTGAGGTTCTCTGAGTGTGCTGCTCAACTGCTCAT 2026
Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
Db 2027 AAGATGTGTACTACGCTTTGGACAGGTTTACACAGAGCAAGCGTCTCTCCAGGCTTT 2086
Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680

Db 2087 GACCGTGCAGAAAGCTGAGATTGGGAATAAAGACATTGAGCTTGATGCTCTGGAGAA 2146
Qy 681 GlyThrThrGluHisThrLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
Db 2147 GCGTATACACAGAACATTGGCTGCTCAGGATATACAGGTAAGGACCTGGGATATCGA 2206
Qy 701 GlyLeuSerArgThr 705
Db 2207 GCGTTGTCAAGGACA 2221

RESULT 2

US-10-028-384-11
; Sequence 11, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM 002219
; DATABASE ENTRY DATE: 2000-10-31
; RELEVANT RESIDUES: (1)..(2472)
US-10-028-384-11

Alignment Scores:

Pred. No.: 0 Length: 2472
Score: 705.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-12 (1-705) x US-10-028-384-11 (1-2472)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 107 ATGACTAAGTTTGGATTTTGGGATTTGCGATTGCTATGAGAGGACACACTTTTGAAGCTT 166
Qy 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGTCATGGCTGCTGATTATCTCTCCACTCGTCTGTTTGTCTGCTGAGA 226
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTGGAAGAGTGTATCCATGAGTTGATCCGTAATTAATTAATTAATTAATTAATTAATTA 286
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTyrTrpProLeu 80
Db 287 GCTGAGAGGGGTTTAAATTCATTAATTCATTAATTCATTAATTCATTAATTCATTAATTC 346
Qy 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGACGAATCATTTGGAGGAACAATTTTACCAGGTTTAAATGATCATCCTGCTGCTGCAATCTAC 406
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTCCACATCACCATCGACATTCCGATGCTGCTGCTGCTGCTGCTGCTG 466
Qy 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CCTCTCTTCT 526

Db 1607 ATATTGATGACTCCGAGAGCATATATTGGCTTCGTCAATAACTCCAGAGGATGCG 1666
 Qy 521 LysValMetSerTTPTrpAspTyrGlyThrAlaMetAlaSerThrIle 540
 Db 1667 AAGGTCATGCTCCTGGTGGATTATGGCTATCAGATTACAGCTATGGCAACCGACAAATT 1726
 Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1727 TTAGTGACAATAACATGGAATAATACCATATTTCTCGAGTAGGGAGGCAATGGCG 1786
 Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1787 TCCACAGAGAAAAGCCTATGAGATCATGAGGAGCTCGATGTACGCTATGTGTGCTC 1846
 Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 Db 1847 ATTTTGGAGGCCCTCACTGGGTATTCCTCTGATGATATCAACAAGTTTCTTTGGATGGTC 1906
 Qy 601 ArgIleGlyGlySerThrAspThrGlyIleHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1907 CGGATTGGAGGAGCAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCA 1966
 Qy 621 ThrGlyGluPheArgValAspArgGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1967 ACTGGGAGTTCCGTGTGGACCGTGAAGTTCTCCAGTGTGCTCAACTGCCTCATGTAC 2026
 Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrAlaLysArgProProGlyPhe 660
 Db 2027 AAGATGTGTACTATCCTTGTGACAGGTTTACACAGAGCAAGCGTCTCCAGGCTTT 2086
 Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 2087 GACCGTGTCCGAATGCTGAGATTGGGAATAAAGACTTTGAGCTGATGCTCTCGAGGAA 2146
 Qy 681 GlyTyrThrThrGluHisTrpLeuValAsnIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2147 GGTATACCAAGAACATTTGGCTGGTCAGGATATACAGGTAAGAGCCTGGATATCGA 2206
 Qy 701 GlyLeuSerArgThr 705
 Db 2207 GGTGTGCAAGGACA 2221

RESULT 3

US-10-172-118-742
 ; Sequence 742, Application US/10172118
 ; Publication No. US20030224374A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dai, Hongyue
 ; APPLICANT: He, Yudong
 ; APPLICANT: Linsley, Peter
 ; APPLICANT: Mac, Mao
 ; APPLICANT: Roberts, Chris
 ; APPLICANT: Van 't Veer, Laura
 ; APPLICANT: Van de Vijver, Marc
 ; APPLICANT: Bernards, Rene
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
 ; FILE OF INVENTION: 9301-175-999
 ; CURRENT APPLICATION NUMBER: US/10/172,118
 ; PRIOR FILING DATE: 2002-06-14
 ; PRIOR APPLICATION NUMBER: 60/380,770
 ; PRIOR FILING DATE: 2002-05-14
 ; NUMBER OF SEQ ID NOS: 2699
 ; SEQ ID NO 742
 ; LENGTH: 2472
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: NM_002219
 ; DATABASE ENTRY DATE: 2001-06-18
 US-10-172-118-742

Alignment Scores:

Pred. No.: 0 Length: 2472
 Score: 705.00 Matches: 705
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 15 Gaps: 0
 US-10-028-384-12 (1-705) x US-10-172-118-742 (1-2472)
 Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 Db 107 ATGACTAAGTTTGGATTTTGGCATTTGCTATGAGAAGCAGGACACACATTTTGAAGCTT 166
 Qy 21 LeuIleLeuSerMetAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 Db 167 CTCATTCTGTCAATGGCTGCTGTATTATCCTTCTCCACTGCTGTGTGCTGCTCCTGAGA 226
 Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 Db 227 TTGGAAGTGTATCCATGAGTTTATCCGTTTATTAATATCGGACTACCGGTTCTCTG 286
 Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspAlaTyrTyrProLeu 80
 Db 287 GCTGAGGAGGGGTTTATAAATTCATTAACCTGGTTTGTATGACCGAGGCTGGTACCCCTTG 346
 Qy 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 347 GGACGAATCATTTGGAGGAGCAATTTACCCAGTTTATGATCACCTCTGCTGCAATCTAC 406
 Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db 407 CATGTACTCCATTTTTCACATCACCATCGCATTCGGAATGCTGTGTGCTGCTGCTGCC 466
 Qy 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 Db 467 CCTCTCTCTCTCTCTCTCCCTCCATCGTCAAGTACCTCTTCCAAAGAGCTCAAGGAT 526
 Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
 Db 527 GCAGGGGCTGGGCTCTTGTGCTGCCATGATGCTGTAGTTCCTGATATATCTCCCGA 586
 Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 Db 587 TCTGTGGCTGGCTCCTATGATAATGAAGGATGCGCATCTTTTGCATGCTACTCACCTAC 646
 Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
 Db 647 TACATGTGGATCAAGGACGATAAAGACTGGTTCCATCTGTGGGAGGAGTAAGTGTGCCCTT 706
 Qy 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIleAsnLeuIlePro 220
 Db 707 GCTTATTCTACATGCTCTCGTCATGGGAGGTTATGTGTTCTCTGATCACTAATTCCT 766
 Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 Db 767 CTCACGCTCTCGTCTGATGCTCAGGCGGCTTTCTCTCACCAGGATCTATGTGGCCTAC 826
 Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db 827 TGTACTGTTTACTGCTGGGTACTATATCTTCTAGGCAGATCTCCTTTGTGGGTTTCCAG 886
 Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 Db 887 CTTGCTCTTTTCATCAGACACATGGCAGGGTTTGGGTCCTTGTGCTCTCTGCCAGATCCAT 946
 Qy 281 AlaPheValAspTyrLysLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db 947 GCCTTTGTGGATTACTTCGCGCAGCAAGTTGAATCCACAACAATTTGAAGTCTCTTTCGG 1006
 Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 Db 1007 AGCGTCATCTCTCTGTGGTGGCTTTTCTCTTCTCACCCTGGGAGCTCTCTCTCATGCTGACA 1066

QY 321 GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db 1067 GGAAGAAATATCTCCCTGGAGGGGGCTTTCTACTCACTGCTGGATCCCTCTTATGCTAAG 1126
QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
Db 1127 AACAAACATCCCATCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1186
QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
Db 1187 TATTTTGACCTGACCTCCTCGCTCTCATGTTTCCAGTTGGCTCTATTACTGCTTAGC 1246
QY 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
Db 1247 AACCTGTCTGATGCGCCGATTTTATCATCATGATGATGATGATGATGATGATGATGAT 1306
QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
Db 1307 GCTGTAAATGGTGGCTTAATGCTAGTGTGGCACCTGTTATGAGCATTCCTCTGGCAT 1366
QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
Db 1367 GGAGTCTCCAGTCTCTCCACATACATGAAGATCTGGACATTAAGTCCGCCAGACAAG 1426
QY 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
Db 1427 AAGACCAAGACCAACAGATTCACCTACCTATTAAGATTGAAGTGGCAAGTGGGATG 1486
QY 461 IleLeuValMetAlaPhePheIleIleThrTyrThrPheHisSerThrTrpValThrSer 480
Db 1487 ATACTGGTCATGGCTTTCTTCTCATCACCTACACCTTTTATCAACCTGGGTGACCACT 1546
QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
Db 1547 GAGGCTACTCTTCTCCGCTCCATGCTACTACTGCTGCTGGGATGGCAGTAGGATC 1606
QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
Db 1607 ATATTGTGATCTCCGAGAGACATATTATTGGCTTCTGATATATCTCAGAGGATGGC 1666
QY 521 LysValMetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
Db 1667 AAGGTCATGCTGCTGGGATATGCTATCAGATACAGCTATGAGCAACCAACCAAT 1726
QY 541 LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1727 TTAGTGGCAATAACACATGGAATAATACCATATTTCTGAGTAGGCGAGCAATGGCG 1786
QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
Db 1787 TCCACAGAGAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTC 1846
QY 581 IlePheGlyGlyIleThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
Db 1847 ATTTTGGAGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1906
QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
Db 1907 CGATTGGAGGAGGACACATACAGGCAACATATCAAGAGATGACTATTATCTCCA 1966
QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1967 ACTGGGAGTTCCTGCTGAGCCGCTGAGGCTTCTCAGTGTCTGCTCAACTGCTCATGTAC 2026
QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
Db 2027 AGATGTGTACTATCTGCTTGGACAGTTTACACAGAGCAACAGCTCTCCAGGCTTT 2086
QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
Db 2087 GACCGTGTCCGAATGCTGAGATTGGGAATAAGACTTTGAGCTTGATGCTCTGGAGGA 2146
QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700

Db 2147 GCCTATACCAAGCAATTCGCTGCTCAGGATATACAGGTAAAGACCTGGATATCGA 2206
QY 701 GlyLeuSerArgThr 705
Db 2207 GCCTTGTCAGGACA 2221
RESULT 4
US-10-342-887-742
; Sequence 742, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; PRIOR FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 16/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-742
Alignment Scores:
Pred. No.: 0 Length: 2472
Score: 705.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0
US-10-028-384-12 (1-705) x US-10-342-887-742 (1-2472)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlnLysGlnAspThrLeuLysLeu 20
Db 107 ATGACTAAGTTTGGATTTTGGGATTCCTATGAGAGCAGGACACACTTTTGAAGCTT 166
QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGTCATGAGGCTGCTGCTATATCTTCTCCACTCGTCTGTTCTGCTGCTGAGA 226
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
Db 227 TTTGAAAGTGTATCCCATGATTGATCCGCTACTTTATATTCGAGTACCAGGTTCTTG 286
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
Db 287 GCTGAGGAGGGGTTTTATAAATTCCTAATCTGTTGATGACCGAGCCTGGTACCCCTTTG 346
QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGACGATCATCTGGAGGACAAATTTACCCAGGTTTATGATCACTCTGCTGCAATCTAC 406
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTTCACATCACCATCGAATTCGGAATCTGCTGTGTTCTCGGCC 466
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140

QY	1	MetThrLysPheGlyPheLeuA:rgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu	20
DB	136	ATGACTAGTTTGGATTATTTTCGGATTGTCTCATGAGACGAGACACACTTTTGAAGCTT	195
QY	21	LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg	40
DB	196	CTCATCTGTCAATGGCTGCTGATTATCTCTCCCACTCGTCGTGTTGCTGCTCCTGAGA	255
QY	41	PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu	60
DB	256	TTTGAAGTGTATCCATGAGTTTGAATCCGTACTTTAAATTATCGGACTACCAAGTTCCTG	315
QY	61	AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTyrProLeu	80
DB	316	GCTGAGGAGGGGTATTAATAATCCATAACTGGTTTATGATCACCGAGCTGCTACCCCTTG	375
QY	81	GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr	100
DB	376	GGACGAATCATTGGAGNACAAATTTACCCAGGTTTAATGATCACTCTGCTGCAATCTAC	435
QY	101	HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla	120
DB	436	CATGTACTCCATTTTTTCCCATCACCATCGAATTCGGAAATGTGTGTGTCTTCCTGGCC	495
QY	121	ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysA	140
DB	496	CCTCTCTTCTCTCTTACCAC-CATCGTCAGGTACCA-CCTTACCAAGAGCTCAGG	553
QY	140	spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerA	160
DB	554	ATGCAGGGGCTGGGCTCTTGTCTGTGCCATGATGCTGTAGTTCTCGATATATCTCCC	613
QY	160	rgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrT	180
DB	614	GATCTGTGGCTGGCTCCTATGATAATGAAGGGATTGCCATCTTTTGCATGCTACTCACT	673
QY	180	yrTyrMetTrpIleValAlaValIleThrGlySerIleCysTrpAlaAlaLysCysAlaL	200
DB	674	ACTCATGTGGATCAAGCAGTAAGACTGGTTCCATCTGTGGGCAGCTAAGTGTGCC	733
QY	200	euAlaTyrPheTyrMetValSerTrpGlyTyrValPheLeuIleAsnLeuIleP	220
DB	734	TGTGCTATTCTACATGGCTCTCGTCATGGGAGGTTATGTGTCTCTGATCAACTTAATTC	793
QY	220	roLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaI	240
DB	794	CTCTCAAGTCTCGTGTGATGTCTCAGCCGCTTCTCTCACCGGATCTATGTGGCCT	853
QY	240	yrCyThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe	259
DB	854	ACTGTACTGTTTACTGCTGGGCATATATCTTCTAT-CCAGATCTCTCTTGTGGGTTC	912
QY	260	GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI	279
DB	913	CAGCCTGTCTTTCATCAGACACATGGCAGC-CTTTGGGGTCTTTGTCTCTGCCAGAT	971
QY	279	ehiAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPhe	299
DB	972	CCATGCTTTTGGGATTACCTGGCAGCAAGTTGAATCCACAACAATTTGAAGTCTCTTT	1031
QY	299	eaRgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLe	319
DB	1032	CCGGAGCGTCATCTCTCTGGTAGCGTTTGTCTTCTCACCGTGGGAGCTCTCTCATGCT	1091
QY	319	unThrGlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAl	339
DB	1092	GAAGAAAAATATCTCCCTGGAGGGGGCTTCTACTCGTGTGGATCCCTCTTAGC	1151
QY	339	abyAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSe	359
DB	1152	TAAGAACAACTCCCACTATGCTCTGTGCTGAGCATCAGCCCAACACTGGTCTC	1211

359	Qy	rTyrThrPheAspLeuGlnIleuValPheMetPheProValGlyLeuTyrTyrCysPh	379
1212	Db	ATACATATTTTACCTCGAGCTCTCGTCTTCATGTTTCCAGTTGGCTCTATTATCTGCT	1271
379	Qy	eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPh	399
1272	Db	TAGCAACCTGTCGTATGCCGGATTTTATCATCATGATATGTTGTGACCAAGATGACTT	1331
399	Qy	eSerAlaValMetValArgIleuMetIleuValIleuAlaProValMetSer-IleIeuSerG	419
1332	Db	TTCACTGTAAATGGTGGCTCTAAATGTAGTGTGGACCTGTTATGTG-CATTCTCTCTG	1390
419	Qy	LYIleuValSerGlnValIleuSerThrTyrMetIysAsnLeuAspIleSerArgProA	439
1391	Db	GCATTGGAGTCTCCAGGTGCTGTCCACATACATGAGAGAACTCGACATAGTGGTCCAG	1450
439	Qy	spLYSerIlySerylGlnIleuSerThrTyrProIleuIlySle-GluValAlaSer	458
1451	Db	ACAAGAAGCAAGCAAGCAAGCAAGATTCACCTTACCTATTAAAGAA-TGAAGTGGCAAGT	1509
459	Qy	GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpVal	478
1510	Db	GGAGTATACTGGTCAATGGCTTTCTTCTCATCCACTACACTTTCATTCAACTGGGTG	1569
479	Qy	ThrSerGluAlaTyrSerSerProSerIleValIleuSerAlaArgGlyGlyAspGlySer	498
1570	Db	ACCAGTGGAGGCTACTCTTCTCGTCCATTGACTATCTGCCCGTGGTGGGATGGCAGT	1639
499	Qy	ArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu	518
1630	Db	AGATCATATTTGATGACTTCCGAGAAGCATATTTGGCTTCGTGCATAAACTCCAGAG	1699
519	Qy	AspAlaLysValMetSerTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg	538
1690	Db	GATCGCAAGGTCATGCTCTGTGGATTATGGCTATCAGATTACAGTATGCGCAACCGA	1749
539	Qy	ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla	558
1750	Db	ACAAATTTAGTGGACATAACACATGGAAATAATACCCATATTTCTCGAGTAGGCGAGCA	1809
559	Qy	MetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal	578
1810	Db	ATGGCGTCCACAGAGGAAAGGCTATGAGATCATGAGGAGCTCGATGTCAGCTATGTG	1869
579	Qy	LeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheIeuTrp	598
1870	Db	CTGGTCATTTTGGAGGCCCTCACTGGGTATTCCTCTGATGATCAACCAAGTTCTTTGG	1929
599	Qy	MetValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyr	618
1930	Db	ATGGTCCGGATTGGAGGGAGCACAGATACAGGCAACATATCAAGAGAGATGACTATTAT	1989
619	Qy	ThrProThrGlyGluPheArgValAspArgGluGlySerProValIleuAsnCysIleu	638
1990	Db	ACTCCAACCTGGGGAGTTCCGTGTGGACCCGTGAAGGTTCTCCAGTCTGCTCAACTGGCTC	2049
639	Qy	MetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProPro	658
2050	Db	ATGTACAGATGTGTACTATCGCTTTGGACAGGTTTACAGAGAGCCAGCGTCTCTCCA	2109
659	Qy	GlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValIleu	678
2110	Db	GGCTTTGACCGGTGCCGAATGTGTGATTTGGGAATAAAGACTTTTGAGCTTGATGTCCTG	2169


```

; APPLICANT: David W. Morris
; APPLICANT: Maic Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 5404
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-99

```

Alignment Scores:	
Pred. No.:	1,61e-227
Score:	226.00
Percent Similarity:	98.25%
Best Local Similarity:	98.25%
Query Match:	23.06%
DB:	18
Gaps:	0
Indels:	12
Mismatches:	6
Conservative:	0
Matches:	674
Length:	5404

US-10-028-384-12 (1-705) x US-10-417-375-99 (1-5404)

1	MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrIleuLysLeu	20
136	ATGACTAAGTTTGGATTTTTCGATATGCTCCTATGAGAAGCAGGACACACTTTTGAAGCTT	195
21	LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg	40
196	CTCATTCCTGCAATGGCTGCTGTATATATCCTTCTCCACTCGTCTGTTTGCCTGCTCAGAA	255
41	PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu	60
256	TTTCAAAAGTGTATCCCATGAGTTGATCCGTACTTTTAATTATCGGACATACCAAGGTTCCTG	315
61	AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu	80
316	CCTGAGGAGGGTTTTATAAATCCATACTGTTTGTATGACGAGCCCTGGTACCCCTTTG	375
81	GlyArgIleLeGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr	100
376	GGACCAATCATGCGAGAACAAATTTACCCAGGTTTAATGATCACCTCTGCTGCAATCTAC	435
101	HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla	120
436	CATGTACTCCCAATTTTTCACATCAACATCGACATTCGGAATGCTGTGTGTCTCGGCC	495
121	ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysA	140
496	CCTCTCTTCTCTCTCTCCCTTCCACCAC-CATCGTCACGTACCA-CCTTACCAAGAGCTCAAGG	553
140	spAlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValValProGlyTyrIleSerA	160
554	ATGCAGGGCGTGGGCTTCTTTCGTCTGCGCAATGCTGTAGTCTCGTGATATATCTCCC	613
160	rgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrT	180
614	GATCTGTGGCTGGCTCCTATGATAATGAAGGATGCCATCTTTTCATGCTACTACCT	673
180	yrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaL	200
674	ACTACATGTGATCAAGGCAGTAAGACATGGTTCCATCTGTTGGGCAGCACTAAGTGTGCC	733
200	euAlaTyrPheTyrMetValSerSerTrpGlyTyrValPheLeuIleAsnLeuIleP	220
734	TTGCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTGTTCTCGATCACTTAATTC	793
220	roLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaIa	240
794	CTCTCCACGCTCTCGTCTGATGCTACAGCGCGTTCTCTCACCGGATCTATGTGGCT	853
240	vrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe	259

Db 1930 ATGGTCGGGATTGGAGGAGCACAGATACAGGCAACATATCAAGGAGAAATGACTATTAT 1989
Qy 619 ThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeu 638
Db 1980 ACTCCAACTGGGAGTTCGGGTGGACCGTGAAGGTTCTCCAGTGTGCTCAACTCCCTC 2049
Qy 639 MetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProPro 658
Db 2050 ATGTACAAAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAGCCAGCGCTCTCCA 2109
Qy 659 GlyPheAspArgValArgAsnAlaGluLeuGlyAsnLysPheGluLeuAspValLeu 678
Db 2110 GCTTTTGACCGTGTCCGAAATGCTGAGATTGGGAATAAGACTTTGAGCTTGATGTCCTG 2169
Qy 679 GluGlu 680
Db 2170 GAGGAA 2175

RESULT 7

US-10-417-375-97
; Sequence 97, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 5827
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-97

Alignment Scores:

Pred. No.: 1,73e-227 Length: 5827
Score: 226.00 Matches: 674
Percent Similarity: 98.25% Conservative: 0
Best Local Similarity: 98.25% Mismatches: 6
Query Match: 32.66% Indels: 12
DB: 18 Gaps: 0

US-10-028-384-12 (1-705) x US-10-417-375-97 (1-5827)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 136 ATGACTAAGTTGGATTTCGCGATTGCTCTATGAGAGGAGGAGACACACTTTTGAGCTT 195
Qy 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 196 CTCATTCTGCAATGGCTGCTGATTATTCCTTCTCCACTGCTGTGTTGCTGCTCGAGA 255
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
Db 256 TTGGAAGTGTATCCATGAGTTTGATCCGTTACTTTAATTATCGGACTACGAGTTCCCTG 315
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
Db 316 GCTGAGGAGGGGTTTATAAATTCATAACTGGTTTATGATGACCGAGCTGTGTTGCTGCTG 375
Qy 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 376 GGACGAAATCATGGAGGAGACAAATTATCCAGGTTTATGATCACCCTCTGCTGCAATCTAC 435
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 436 CATGACTCCATTTTTCACATCACCACATCGACATTCGGAATGCTGTGTGTTCTCCCTGGCC 495
Qy 121 ProLeuPheSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysA 140

Db 496 CCT 553
Qy 140 spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
Db 554 ATGACAGGGCTGGGCTTCITGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 613
Qy 160 rgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 180
Db 614 GATCTGTGGCTGGCTCTCTATGATAATGAAGGAGTTCCTCTCTCTCTCTCTCTCTCTCTCT 673
Qy 180 yTyrMetTyrIleLysAlaValLysThrClySerIleCysTrpAlaAlaLysCysAla 200
Db 674 ACTACAGTGAATCAAGGAGTAAAGACTGCTTCCATCTGTTGGGCGAGTAAAGTGGCC 733
Qy 200 euAlaTyrPheTyrMetValSerSerTrpClyGlyTyrValPheLeuLeuLeuLeuP 220
Db 734 TTGCTTATTTCTACATGCT 793
Qy 220 xLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAla 240
Db 794 CTCTCCACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 853
Qy 240 yTyrThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
Db 854 ACTGCTACTGTTTACTGCTGGGCACTATCTTTCTAT-GCAGATCTCTCTCTCTCTCTCTCTCT 912
Qy 260 GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGln 279
Db 913 CAGCTGCTCTCTCTCATCAGAGCACATGGCAGC-CTTTGGGGCTTTTGGTCTCTGCCAGAT 971
Qy 279 eHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPh 299
Db 972 CCATGCTTTGTGGATTACCTGGCGCAGCAAGTGAATCCACAACTTTGAAGTCTCTTT 1031
Qy 299 eArgSerValIleSerLeuValGlyPheValLeuThrValGlyAlaLeuMetIle 319
Db 1032 CCGAGCGCTCATCTCTGTTAGGCTTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1091
Qy 319 uThrGlyLysSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAl 339
Db 1092 GACAGGAAAAATATCTCTCTGGAGGGGGCTTTCTACTCGCTGCTGCTGCTCTCTCTCTCTCT 1151
Qy 339 aLysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrSerSe 359
Db 1152 TAAGAACAAATCCCCATCATCTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1211
Qy 359 rTyrTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPh 379
Db 1212 ATACTATTTTGACCTGCGAGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1271
Qy 379 eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPh 399
Db 1272 TAGCAACCTGCTGATGCCGGATTTTATCATCATGATATGTTGTTGACGAGCATGACTT 1331
Qy 399 eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG 419
Db 1332 TTCAGCTGAATGTTGCTGCTAATGCTAGTGTGGCACCCTGTTATGTG-CATTCTCTCTG 1390
Qy 419 lYleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProA 439
Db 1391 GCATTGGAGTCTCCAGGCTGCTGCACATACATGAAGAACTCGGACATAAGTCGTCAG 1450
Qy 439 spLysLysSerLysLysGlnAspSerThrTyrProIleLysIle-GluValAlaSer 458
Db 1451 ACAAGAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 1509
Qy 459 GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpVal 478
Db 1510 GGGATGATACATGGTATGGCTTTCTCTCTCATCCTTACCTTCTCATCAACCTGGGTG 1569
Qy 479 ThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer 498
Db 1570 ACCAGTGAGGCTACTCTCTCTCGCTCCATGTTACTATCTCTCTCTCTCTCTCTCTCTCTCT 1629

```

QY 499 ArgillePheAspAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProGlu 518
Db 1630 AGGATCATATTGATGACTCCGAGAGCATATATTGGCTCGTCATATACTCCAGAG 1689

QY 519 AspAlaValMetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg 538
Db 1690 GATCGAAGGTCAATGCTGCTGGATTATGGCTATCAGATTACAGCTATGGCAACCGA 1749

QY 539 ThrIleLeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAla 558
Db 1750 ACAATTTAGTGACATAACACATGGAATAATACCCATATTTCTCGAGTAGGCGAGCA 1809

QY 559 MetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal 578
Db 1810 ATGGCTCCACAGAGGAAAAGGCTATGAGATCATGAGGAGCTCGATGTCAGCTATGTG 1869

QY 579 LeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTyr 598
Db 1870 CTGGTCATTTTGGAGGCTCACTGGGTATTTCTCTGATGATATCAACAAGTTTCTTTGG 1929

QY 599 MetValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyr 618
Db 1930 ATGGCTCCGATGGAGGAGCACAGATACAGGCAACATATCAAGAGATGACATATAT 1989

QY 619 ThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeu 638
Db 1990 ACTCCAACTGGGAGTTCCGTGTGGACCGTGAAGGTTCCTCCAGTGTGCTCAACTGCCTC 2049

QY 639 MetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProPro 658
Db 2050 ATGACAGATGCTACTACTCTTTGGACAGTTTACAGAGTTTACACAGAGCCAGGCTCTCA 2109

QY 659 GlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeu 678
Db 2110 GGCTTTGACCGTGTCCGAAATGCTGAGATGGGAATAAAGACTTTGAGCTTGATGCTGTG 2169

QY 679 GluGlu 680
Db 2170 GAGGAA 2175

RESULT 8
US-10-417-375-92
; Sequence 92, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 3093
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-417-375-92

Alignment Scores:
Pred. No.: 1,36e-149 Length: 3093
Score: 152.00 Matches: 667
Percent Similarity: 97.37% Conservative: 0
Best Local Similarity: 97.37% Mismatches: 9
Query Match: 21.56% Indels: 18
DB: 18 Gaps: 0

US-10-028-384-12 (1-705) x US-10-417-375-92 (1-3093)

QY 5 GlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeuLeuLeuSer 24
Db 124 GGATTTTGGATTGCTTCAATGAGAGGAGGACACACTTCTAAAGCTTCTCATCTCTGTCG 183

```

```

QY 25 MetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerVal 44
Db 184 ATGGCTGCTGTGTATCTTTTCTACTCGTCTTTTGTGCTGTGCTGAGATTGAAAGTGTC 243

QY 45 IleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGluGly 64
Db 244 ATCCATGAGTTTGATCCCGTACTTTTAATTAATCGGACTACCCCGTTTCTGGCTGAGGAGGG 303

QY 65 PheTyrLysPheHisAsnTyrPheAspArgAlaTyrTyrProLeuGlyArgIleIle 84
Db 304 TTTTATAATTCATTAACCTGGTTTGATGACCGGGCTTGGTACCCCTTTGGGCCCAATCAT 363

QY 85 GlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyrHisValLeuHis 104
Db 364 GAGGACAAATTTACCCAGGTTTATGATCACTTCTGCTGCATCTACCAATGACTCCAT 423

QY 105 PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSer 124
Db 424 TTTCTCCATATCACCTATTGACATTCGGAATGCTGTGTGTTTCTCTGGGCCCACTTTTCTCC 483

QY 125 SerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysAspAlaGlyAlaG 144
Db 484 TCTTTCACAC-CATCGTTAGTACCA-CCTTACCAAGAGCTCAAGATGCGAGAGCTG 541

QY 144 LysLeuAlaAlaAlaMetIleAlaValProGlyTyrIleSerArgSerValAlaG 164
Db 542 GGCTTCTGCTGCTGCCATGATGCTGTAGTTCCTGGGTATATTTCTCGATCTGTAGCTG 601

QY 164 LysTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetTrpI 184
Db 602 GTCTCTATGATAATGAAGAAATGCTATCTTTGCTGCTGCTTACTTACTTACTCATGTGGA 661

QY 184 LysAlaValLysThrGlySerIleCys-TrpAlaAlaLysCysAlaLeuAlaTyrPhe 203
Db 662 TCAAGCGAGTGAAGACTGGTTCATCTA-TTGGGCTGCCAAGTGTGCCCTCGCTTATTTC 720

QY 204 TyrMetValSerSerTrpGlyGlyTyrValPheLeuIleLeuLeuLeuProLeuHisVal 223
Db 721 TACATGGTCTCTTCAATGGGAGGCTATGTTCTCTGATCACTTGTATCTCTACATGTC 780

QY 224 LeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrVal 243
Db 781 CTGGTGTCTAATGCTACAGAGCGCTTTTCTCACCGATCTACGTAGCTACTGTACTGTT 840

QY 244 TyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPheGlnProValle 263
Db 841 TACTGCTGGGACCATCTCTTCTAT-CGAGATTCTCTTTGTTGGTTTCCAGCCCGTCT 899

QY 263 uSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnIleHisAlaPheV 283
Db 900 TTTATCAGAACACATGGCAGG-CATTGGAGTGTTTGGTCTCTGTGATCATCATGCTTTCG 958

QY 283 AlaSerTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArgSerValI 303
Db 959 TGAATTTACTGCGCAGCAAGTTGAATCCACAGCAATTCGAAGTTCTTTTCCGAGAGTTA 1018

QY 303 LysLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThrGlyLysI 323
Db 1019 TCTCCCTGTTGGCTTTGCTCTCTCTACTGTGGAGGCTCTCTCATCTAACAGAAAAA 1078

QY 323 LysSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLysAsnAsnI 343
Db 1079 TTTTCTCCCTGGACAGGCGTTTCTACTCTCTGCTGGATTCCTCTTATGCTTAAAGATAACA 1138

QY 343 LysProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyrTyrPheA 363
Db 1139 TTTCCCATTTATGCACTCTGTTTCTGAGCACCAGCCCAACAGCTGCTCTCTTACTATT 1198

QY 363 sLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSerAsnLeuS 383
Db 1199 ATCTACAGCTCTCTGCTCTTCAATGTTTCAAGTTGGCTCTTATCTGCTTTAGCACTGT 1258

```

```

QY 383 exAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSerAlaValM 403
Db 1259 CTGATGCTCGGATTTTATCATCATGATGATGATGATGATGATGATGATGATG 1318
QY 403 etValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerGlyIleGlyVal 422
Db 1319 TGGTCGGCTAATGCTGATATGGCACCCTGTTATGTG-CATCTCTTCTGCGATGTTGTT 1377
QY 423 SerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLysLysSer 442
Db 1378 TCCAGGTCCTGCTCACATATATGAANAATCTGGACATAGTCGCCACAGACAGAGC 1437
QY 443 LysLysGlnGlnAspSerThrTyrProIleLysIle-GluValAlaSerGlyMetIleLe 462
Db 1438 AAGAAGCAACAGGATTCATCTACCTATTAAGAA-TGAGGTGGCGAGTGGGATGATCT 1496
QY 462 uValMetAlaPhePheLeuIleThrTyrPheHisSerThrTyrValThrSerGluAl 482
Db 1497 GGTCAATGCTTTTCTCATCACTACCTACACGCTTCATTCGACTTGGGTGACCGAGTGA 1556
QY 482 atYrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleLeP 502
Db 1557 CTATTCCTTCTCCCTCATGATCTGCTGCTGCTGGGATGGCAGTAGTAGGATCATTTT 1616
QY 502 eAspAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProGluAspAlaLysVa 522
Db 1617 TGATGACTTCCAGAGAGGATTTATTTGCTCCCTCACAATACTCCAGAGATGCCAAAGT 1676
QY 522 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVa 542
Db 1677 CATGTCATGGTGGGATTTATGCTACCAATTTACTGCAATGGCAATCGGCAATTTTAGT 1736
QY 542 tAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerTh 562
Db 1737 GGACATTAACATGGAATATATCCATATTTCTCGAGTAGGGCAGCAATGGCATCCAC 1796
QY 562 rGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePh 582
Db 1797 AGAAGAAAAGCCTATGAATCATGAGGAGCTTGTATGTCAGCTATGCTTGTCTATTTT 1856
QY 582 eGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArgIl 602
Db 1857 TGGAGGCTTACTGGTATTTCTTGGATGATATCACAGTTCTTTGGATGTCGGAT 1916
QY 602 eGlyGlySerThrAsp-ThrGlyLys-HisIleLysGluAsnAspTyrTyrTrpThr 621
Db 1917 TGGAGGAAGACACAGA-GACAGG-AAGACACATTAAAGGAGATGACTACTATCTCTACT 1974
QY 622 GlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLys 641
Db 1975 GGGGAATTCGGTGTATCGTAGGGTTCTCCGTGCTGCTCAACTGCCATTATGTACAA 2034
QY 642 MetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAsp 661
Db 2035 ATGTGTACTACCGCTTTGGGCGAGTCTACACAGAGCCAAAGCGTCCACCGAGCTTTGAC 2094
QY 662 ArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
Db 2095 CGTGTTCGAATGCTGAGATTTGTAATAAGACTTTGAGCTTGTATGCTTGGAGAA 2151

```

RESULT 9

```

US-10-028-384-9
; Sequence 9, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13

```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 3094
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_008408
; DATABASE ENTRY DATE: 2000-11-01
; RELEVANT RESIDUES: (1)..(3094)
US-10-028-384-9

Alignment Scores:
Pred. No.: 1,366-149 Length: 3094
Score: 152.00 Matches: 667
Best Similarity: 97.37% Conservative: 0
Best Local Similarity: 97.37% Mismatches: 9
Query Match: 21.56% Indels: 18
DB: 15 Gaps: 0

US-10-028-384-12 (1-705) x US-10-028-384-9 (1-3094)

```

```

QY 5 GlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeuLeuLeuSer 24
Db 124 GGATTTTGGCATTTGCTATGAGAGACAGACACACTTCTAAAGCTTCTCATCTGTCG 183
QY 25 MetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerVal 44
Db 184 ATGGCTGCTGCTGATCTTTTCTTACTCGCTTTTCTGCTGCTGAGATTGAAAGTGC 243
QY 45 IleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGluGly 64
Db 244 ATCCATGAGTTGATCGCTACTTTTAAATATCGGACTACCCCGTTTCTGGCTGAGGAGGG 303
QY 65 PheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeuGlyArgIleIle 84
Db 304 TTTTATAAATTCCTACTACTGTTTGGTACCGGCTTGGTACCTTTGGCCCAATCAIT 363
QY 85 GlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHis 104
Db 364 GGAGAAACAAATTTACCAGGTTTAAATGATCATCTTCTGCTGCAATCTACCATGCTCAT 423
QY 105 PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSer 124
Db 424 TCTTTCCATATCATCTATTTGCAATTCGGAATGCTGTGTTTCTGGCCCACTTTTCTCC 483
QY 125 SerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysAspAlaGlyAlaG 144
Db 484 TCTTTTACCAC-CATCGTTACGTACCA-CCTTACCAAGAGAGCTCAAGGATGAGGAGCTG 541
QY 144 LysLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArgSerValAlaG 164
Db 542 GGCCTTCTTGGCTGCGCATGATGCTGTATTTCTGGGTATATTTCTCGATCTGTAGCTG 601
QY 164 LysTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetTrpI 184
Db 602 GCTCTATGATAAAGAGGAATGCTATCTTTTGATGCTGCTTACTTACTACATGGA 661
QY 184 IleLysAlaValLysThrGlySerIleCys-TrpAlaAlaLysCysAlaLeuAlaTyrPhe 203
Db 662 TCAAGGCACTGAAGACTGGTTCATCTA-TTGGGCTGCCAAGTGTGCCCTCGCTTATTC 720
QY 204 TyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuIleProLeuHisVal 223
Db 721 TACATGGTCTCTTTCATGGGAGGCTATGCTTCTGATCACTTGAATCTCTTACATGTC 780
QY 224 LeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrVal 243
Db 781 CTGTGTCTAATGCTGACAGCCCGTTTTTCTCACCGGATCTACGTAGCTACTTACTGT 840
QY 244 TyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPheGlnProValLe 263
Db 841 TACTGCTGGGACCACTTCTCTAT-GCAGATTCTCTTTGTTGGTTTCCAGGCCCTCTCT 899

```



```

Db      19 CTGCTCAACTGCCTCATG 2
RESULT 11
US-09-904-456-59/c
; Sequence 59, Application US/09904456
; Publication No. US20030017167A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.524C1
; CURRENT APPLICATION NUMBER: US/09/904,456
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 247
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-904-456-59
Alignment Scores:
Pred. No.:      4,566-144      Length:      440
Score:          146.00        Matches:      146
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    20.71%        Indels:      0
DB:             10           Gaps:         0

US-10-028-384-12 (1-705) x US-09-904-456-59 (1-440)
QY      494 GlyGlyAspGlySerArgIlePheAspAspPheArgGluAlaTyrTyrTyrLeuArg 513
Db      439 GGTGGGATGGCAGTAGATCATATTTGATGACTCCGAGAAGCATATATTGGCTTCGT 380
QY      514 HisAsnThrProGluAspAlaLysValMetSerTrpAspTyrGlyTyrGlnIleThr 533
Db      379 CATATATCTCCAGAGATCCGAGGTCTATGTCCTGGTGGGATATGCTATCATCATACA 320
QY      534 AlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSer 553
Db      319 GCTATGGCAACCCGAACAATTTAGTGGACAATAACACATGGAATATATCCCATATTTCT 260
QY      554 ArgValGlyGlnAlaMetAlaSerThrCluLysAlaTyrGluIleMetArgGluLeu 573
Db      259 CGAGTAGGGCAGCCGATGGGTCCACAGAGGAAAAGCCATAGATCATCAGGGAGCTC 200
QY      574 AspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIle 593
Db      199 GATGTCAGCTATGCTGCTCATTTTGGAGGCTCCTGCTGATGATC 140
QY      594 AsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHisIleLys 613
Db      139 AACAAATTTCTTTGGATGGTCCGATGGAGGAGCAGATACAGCAACATATCAAG 80
QY      614 GluAsnAspTyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerProVal 633
Db      79 GAGAATGACTATATATCTCCAACTGGGAGTTCGGTGGACCGTGGAGGTTCTCCAGTG 20
QY      634 LeuLeuAsnCysLeuMet 639
Db      19 CTGCTCAACTGCCTCATG 2

RESULT 12
US-09-878-178-1456
; Sequence 1456, Application US/09878178
; Patent No. US2003017552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

```

```

; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1456
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1456
Alignment Scores:
Pred. No.:      1,02e-137      Length:      452
Score:          140.00        Matches:      140
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    19.86%        Indels:      0
DB:             9           Gaps:         0

US-10-028-384-12 (1-705) x US-09-878-178-1456 (1-452)
QY      490 LeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyr 509
Db      2 CTATCTGCCCGTGGTGGGATGGCAGTAGGATCATATTTGATGACTTCGAGAAGCATAT 61
QY      510 TyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpAspTyrGly 529
Db      62 TATTGGCTTCGTATATACTCCAGAGGATGCCAAGGTATGCTCTGGTGGGATATGGC 121
QY      530 TyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsn 549
Db      122 TATCAGATTACAGCTATGGCAACCCGAACAATTTAGTGACAATAAACACATGGAATAAT 181
QY      550 ThrHisIleSerArgValGlyGlnAlaMetAlaSerThrCluLysAlaTyrGluIle 569
Db      182 ACCCATATTTCTCAGTAGGGCAGGCAATGGCTCCACAGAGGAAAAGCCATAGATC 241
QY      570 MetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSer 589
Db      242 ATGAGGGAGCTCGATGTCAGTATGCTGTCATTTTGGAGGCTCACTGGGTATTCC 301
QY      590 SerAspAspIleAsnLysPheLeuTrpMetValArgIleGlySerThrAspThrGly 609
Db      302 TCTGATGATATCAACAAGTTTCTTTGGATGGTCCGATTTGGAGGAGCAGATACAGC 361
QY      610 LysHisIleLysGluAsnAspTyrTyrThrProThrGlyGluPheArgValAspArgGlu 629
Db      362 AACATATCAAGGAGATGACTATTATCTCCAACTGGGAGTTCCGTGTGGACCGTGAA 421

RESULT 13
US-10-046-935-1456
; Sequence 1456, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aljun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1456
; LENGTH: 452

```

TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 422
 OTHER INFORMATION: n = A,T,C or G
 US-10-046-935-1456

Alignment Scores:

Pred. No.: 1,028-137 Length: 452
 Score: 140.00 Matches: 140
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 19.86% Indels: 0
 DB: 13 Gaps: 0

US-10-028-384-12 (1-705) x US-10-046-935-1456 (1-452)

QY 490 LeuSerAlaArgGlyGlyAspGlySerArgGlyLeuVallePheAspAspPheArgGluAlaTyr 509
 DB 2 CTATCTCCCGTGGTGGGATGGCAGTAGGATCATATTTGATGACTTCCGAGAAGCATAT 61
 QY 510 TyrTrpLeuArgHisAsnThrProGluAspAlaValMetSerTrpTrpAspTyrGly 529
 DB 62 TATTGGCTTCGTATATATCTCCAGAGGATGCGAAGGTCTATGCTCTGGTGGGATTATGCG 121
 QY 530 TyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsn 549
 DB 122 TATCAGATTACAGCTATGGCAACCCGAAATTTTAGTGACAATAAACACATGGGAATAAT 181
 QY 550 ThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIle 569
 DB 182 ACCATATTTCTCGAGTAGGGAGGCAATGGCGTCCACAGAGGAAAAAGCCCTATGAGATC 241
 QY 570 MetArgGluLeuAspValSerTyrValLeuVallePheGlyGlyLeuThrGlyTyrSer 589
 DB 242 ATGAGGAGCTCGATGTCAGCTATGCTGTGTCATTTTGGAGGCTCCTACTGGGTATTC 301
 QY 590 SerAspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThrAspThrGly 609
 DB 302 TCTGATGATATCAACAGTTTCTTTGGATGTCGCGATGGAGGAGCACAGATACAGGC 361
 QY 610 LysHisIleLysGluAsnAspTyrTyrProThrGlyGluPheArgValAspArgGlu 629
 DB 362 AAACATATCAAGGAGATGACTATTATCTCAACTGGGGAGTTCGCTGTGGACCGTGAA 421

RESULT 14

US-10-146-502-1456
 Sequence 1456, Application US/10146502
 Publication No. US20030069180A1
 GENERAL INFORMATION:
 APPLICANT: Jiang, Yugu
 APPLICANT: Harlocker, Susan L.
 APPLICANT: Secrist, Heather
 APPLICANT: Wang, Aijun
 APPLICANT: Stolk, John A.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 FILE REFERENCE: 210121.527C2
 CURRENT APPLICATION NUMBER: US/10/146,502
 CURRENT FILING DATE: 2002-05-14
 NUMBER OF SEQ ID NOS: 2241
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 1456
 LENGTH: 452
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 422
 OTHER INFORMATION: n = A,T,C or G
 US-10-146-502-1456

Alignment Scores:

Pred. No.: 1,028-137 Length: 452
 Score: 140.00 Matches: 140
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 19.86% Indels: 0
 DB: 14 Gaps: 0

US-10-028-384-12 (1-705) x US-10-146-502-1456 (1-452)

QY 490 LeuSerAlaArgGlyGlyAspGlySerArgGlyLeuVallePheAspAspPheArgGluAlaTyr 509
 DB 2 CTATCTCCCGTGGTGGGATGGCAGTAGGATCATATTTGATGACTTCCGAGAAGCATAT 61
 QY 510 TyrTrpLeuArgHisAsnThrProGluAspAlaValMetSerTrpTrpAspTyrGly 529
 DB 62 TATTGGCTTCGTATATATCTCCAGAGGATGCGAAGGTCTATGCTCTGGTGGGATTATGCG 121
 QY 530 TyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsn 549
 DB 122 TATCAGATTACAGCTATGGCAACCCGAAATTTTAGTGACAATAAACACATGGGAATAAT 181
 QY 550 ThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIle 569
 DB 182 ACCATATTTCTCGAGTAGGGAGGCAATGGCGTCCACAGAGGAAAAAGCCCTATGAGATC 241
 QY 570 MetArgGluLeuAspValSerTyrValLeuVallePheGlyGlyLeuThrGlyTyrSer 589
 DB 242 ATGAGGAGCTCGATGTCAGCTATGCTGTGTCATTTTGGAGGCTCCTACTGGGTATTC 301
 QY 590 SerAspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThrAspThrGly 609
 DB 302 TCTGATGATATCAACAGTTTCTTTGGATGTCGCGATGGAGGAGCACAGATACAGGC 361
 QY 610 LysHisIleLysGluAsnAspTyrTyrProThrGlyGluPheArgValAspArgGlu 629
 DB 362 AAACATATCAAGGAGATGACTATTATCTCAACTGGGGAGTTCGCTGTGGACCGTGAA 421

RESULT 15

US-09-871-161-213
 Sequence 213, Application US/09871161
 Publication No. US20030097666A1
 GENERAL INFORMATION:
 APPLICANT: ENDEGE, WILSON O., ET AL.
 TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
 FILE REFERENCE: CCNDA-260XX
 CURRENT APPLICATION NUMBER: US/09/871,161
 CURRENT FILING DATE: 2001-05-31
 PRIOR APPLICATION NUMBER: 09/328,111
 PRIOR FILING DATE: 1999-06-08
 PRIOR APPLICATION NUMBER: 60/117,393
 PRIOR FILING DATE: 1999-01-27
 PRIOR APPLICATION NUMBER: 60/098,639
 PRIOR FILING DATE: 1998-08-31
 NUMBER OF SEQ ID NOS: 544
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 213
 LENGTH: 487
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-871-161-213

Alignment Scores:

Pred. No.: 5,168-125 Length: 487
 Score: 128.00 Matches: 128
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 18.16% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-12 (1-705) x US-09-871-161-213 (1-487)

```
QY 271 PheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeu 290
Db      |||||
QY 291 AsnProGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPheValLeu 310
Db      |||||
QY 311 LeuThrValGlyAlaLeuLeuMetLeuThrGlyLysIleSerProTyrThrGlyArgPhe 330
Db      |||||
QY 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleIleAlaSerValSer 350
Db      |||||
QY 351 GluHisGlnProThrThrTyrSerSerTyrTyrPheAspLeuGlnLeuValPheMet 370
Db      |||||
QY 371 PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleIle 390
Db      |||||
QY 391 MetTyrGlyValThrSerMetTyr 398
Db      |||||
QY 448 ATGTATGGTGTGACCATGTAC 471
```

Search completed: December 15, 2004, 15:28:24
Job time : 930.605 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:58 ; Search time 5457.4 Seconds
(without alignments)
4707.376 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 705

Sequence: 1 MTKGFLRLSVERQDTLLKL.....HMLVRIYKYKDLNRLSLST 705

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-O=/cgn2.1/USP20.spool/US10028384/runat.14122004.131646.6807/app_query.fasta.1.3740
-DB=EST -OFMT=FASTAP -SUFFIX=Oligo.rst -MINMATCH=0.1 -LOOPTC=0 -LOOPTXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn.1.13960 @runat.14122004.131646.6807 -NCPU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FSAPOP=6
-FSAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hcc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	226	32.1	1345	3	CR606556	CR606556 full-length
2	218	30.9	1145	5	BX446691	BX446691 BX446691
3	210	29.8	845	4	BG678139	BG678139 602625911
4	198	28.1	2118	9	AV418284	AV418284 Homo sapi
5	183	28.0	770	4	B1859687	B1859687 603386275
6	179	25.4	686	1	AU123167	AU123167 AU123167
7	178	25.2	541	7	CN265339	CN265339 170005999
8	178	25.2	692	1	AI905439	AI905439 RC-BT091-
9	164	23.3	994	6	CA487523	CA487523 AGENCOURT

10	161	22.8	484	5	BX284031	BX284031
11	160	22.7	733	7	CK38051	CK38051 4063016 B
12	157	22.3	914	6	CA454755	CA454755 AGENCOURT
13	155	22.0	615	1	AI437149	AI437149 fb38c11.Y
14	155	22.0	627	2	BE251342	BE251342 601107886
15	155	22.0	743	5	BU471816	BU471816 603760287
16	155	22.0	760	5	BU248580	BU248580 603779426
17	155	22.0	776	7	CK027267	CK027267 AGENCOURT
18	155	22.0	816	5	BU200678	BU200678 603101974
19	155	22.0	824	5	CK028993	CK028993 AGENCOURT
20	155	22.0	987	5	BU356020	BU356020 603474684
21	155	22.0	1124	7	CK030365	CK030365 AGENCOURT
22	154	21.8	927	5	BU543738	BU543738 AGENCOURT
23	153	21.7	556	6	CB363245	CB363245 ZF001-P00
24	152	21.6	639	6	CA531158	CA531158 C0315G06-
25	152	21.6	694	4	B1654677	B1654677 603285595
26	152	21.6	935	6	CB208848	CB208848 AGENCOURT
27	152	21.6	2103	9	AY418286	AY418286 Mus muscu
28	152	21.6	2656	3	BC028897	BC028897 Mus muscu
29	152	21.6	2730	3	AK081547	AK081547 Mus muscu
30	152	21.6	3802	3	AK030363	AK030363 Mus muscu
31	151	21.4	643	5	BU290292	BU290292 604164809
32	148	21.0	481	7	CF931151	CF931151 CF--06-R-
33	147	20.9	576	4	BMS11440	BMS11440 J48C04.Y
34	147	20.9	814	4	B1684137	B1684137 603306543
35	146	20.7	620	6	CB426082	CB426082 601223 MA
36	145	20.6	465	1	AJ692414	AJ692414 AJ692414
37	145	20.6	626	5	BQ526963	BQ526963 NISC no19
38	145	20.6	641	1	AL865732	AL865732 AL865732
39	145	20.6	849	7	CR409802	CR409802 CR409802
40	145	20.6	860	7	CR416594	CR416594 CR416594
41	145	20.6	864	7	CO788779	CO788779 spleen_ES
42	145	20.6	941	7	CR585625	CR585625 CR585625
43	144	20.4	600	6	CA527576	CA527576 8043-93 M
44	144	20.4	650	7	CN265342	CN265342 170005327
45	144	20.4	688	2	BF348828	BF348828 QV0-DT002

ALIGNMENTS

CR606556 1345 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CLOB0092D10 of Neuroblastoma of Homo sapiens (human).

ACCESSION CR606556
VERSION - GI:50487363
KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
2 (bases 1 to 1345)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

REFERENCE 2 (bases 1 to 1345)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES
source
Location/Qualifiers
1..1345
/organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="Cl0BB009ZD10"
/tissue_type="Neuroblastoma"
/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:
Pred. No.: 1,87e-232 Length: 1345
Score: 226.00 Matches: 336
Percent Similarity: 98.82% Conservative: 0
Best Local Similarity: 98.82% Mismatches: 2
Query Match: 32.06% Indels: 4
DB: 3 Gaps: 0

```

Accession	Protein	Score	Length	Ident	Positives	Negatives
662	rgvAlrGAsnAlaGluIleGlyAsnLysAspPheGluIeuAspValIeuGluGlu 680	100	100	100	100	100
959	GTGTCCGAATGTCTGAGATTGGGAATAAGACTTTGAGCTTGATGCTCTGAGGAA 1014	100	100	100	100	100

US-10-028-384-12 (1-705) x CB606556 (1-1345)

QY 343 IleProIleileAlaSerValSerGluHisGlnProThrThrTrpSerSerTyrrPhe 362

1 ATCCCCATCATGCTTCTGTGTGTGAGCATCAGCCACACCTGGTCTCATACTATTTT 60

363 AspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrCysPheSerAsnLeu 382

Db
61 GACCTGCAGTCTCTGTCTTCATGTTTCCAGTTGGCTCTATTACTGTTTAGCAACCTG 120

Qy 383 SerAspAlaArgGilePheIleileMetTyrGlyValThrSerMetTyrPheSerAlaVal 402

db
121 TC¹GATGCCGGATTTATCATCATGTGTTGACCCAGCATGTACTTTTCAGCTGA 180

Qy 403 MetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerGlyIleGlyVa 422

Db
181 ATGGTGGCTAATGCTAGTGTGGCACCTGTTATGTG-CATTCTCTGGCATTGGAGT 239

Qy 422 1serGlnValLeuSerThrTyMetLysAsnLeuAspIleSerArgProAspLysLysse 442

Db
240 CTCCGAGGTGTCCACATACATGAAGAATCTGGACATRACTGTCCAGACAGAAGAG 299

DB	240	CTCCAGGTGTGTCTCCACATACATACATGAAGAATCTGGACATAACTCGTCCGACACAAGAAGAG	462
QY	442	rlvslvsglgnlgnAspserThrtvrvProilelvslile-GluValAlaserGlvwtilel	462

QY 442 rLysLysGlnGlnAspSerThrThrProileLysile-GluValAlaSerGlyMetIleL 462

Db 300 CAAGAAGCAACAGGATTCACCTACCTATTAGAA-TGAAGTGGCAAGTGGGATGATAC 358

DB 300 CAAGAAGCAACAGGATTCACACCTACCTATTAGAA-TGAAGTGGCAAGTGGGATGATAC 358

OV 452 enlValMetAlldheBheLeuTlTethrTyrThrDhdh;eSerThrTrnValhSerClnA 482

QY 462 euValMetAlaPhePheLeuIleThrTyThrPheHisSerThrTrpValThrSerGluA 482

482 euvalmetcalapnepheleuineirlyinipnhesiseriniripvalinserrina 488
359 TGGTCATGGCTTTCTTTCTCATCACCTACACCTTTTCATTCACCTGGGTGACCACTGAGG 418

Db 359 TGGTCATGGCTTTCTTTCTCATCCCTACCTTCACTTCACTGGGTGACAGTGAGG 418

Qy 482 laTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIleP 502

Qy	482	LatYrSerSerProSerIleValLeuSerAlaAlaArglyGlyVasgylVserArgIleIleIle	506
Db	419	CCTACTCTTCTCCGTCGCAATGTACTATCTGCCCGGTGGGATGGCAATCATAT	478

Db 419 CCTACTCTTCTCCGTCCATTGTACTATCTGCCCGTGGGATGGCAGTAGGATCATAT 478

QY 502 heAspAsp pheArgGlu alaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysV 522

502	heAspAspPheArgGluAlaTyrTyrTrrPleuArgHisAsnThrProGluAspAlaLysV	522
Qy		
Db		
479	TTGATGACATCCGAGAGACATATTATGGTTTCGTCATAAATCTCCAGAGATCCGAAGG	538

Db 479 TTGATGACITCCGAGAAGCATATTATTGGCTTCGTCATAAIACTCCAGAGGATCGAAGG 538

522 **Qy** **al** **WetSer** **Trp** **Trp** **Asp** **Tyr** **Gly** **Tyr** **Gln** **ile** **Thr** **Ala** **Met** **Ala** **Asn** **Arg** **Thr** **ile** **Leu** **542**

Qy 522 alVetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuV 542
Dh 539 TCATGTCTTCAGCTGGCATTTATTCAGTATATCAATTATTCAGCTATACCCATACCGCAGATTTTTC 598

Db 539 TCA TGTCCTCGTGGGATATGGCTATCAGATTACAGCTATGCAACCGACAAATTTTAG 598

LD	339	ICATGICCCGGGGGATATGGCTATCAGATTACACTAATGCAAAACCGAACAAATTTTAG	598
QY	542	alaspAsnAsnThrTTPasnAsnThrHisIleSerArgValGlyGlnIleAlaSerT	562

542 **QY** `alaaspAsnAsnThrTrrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerT` 562

Db
599 TGGACAATAACATGGAATAATACCCATATTTCTGAGTAGGGCAGGCAATGGCGTCCA 658

DB	599	TGGACAATAACACATCGGAATAATACCCATAATTTCTCGAGTAGGGCAGGCAATGCGGTCCA	658
QY	562	hrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValIleValIleP	582

Qy 562 hrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIleP 582

Db
659 CAGAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGCAGCTATGCTGCTGCTCATTT 718

	Dbb	CAGAGGAAAAAAGCCTATGAGATCATGTGGAGGACTCGATGTCAGCTATGTGCTGGTCAATT	719
	Ov	beglvglvleuthrqlgvtrvsrraspsapilcaasnLvaphelcetrMetValArat	602

QY 582 heGlyGlyLeuThrGlyTyrSerSerAspPileasnLysPheLeuTrpMetValArgi 602

Db 719 TTGAGGCCCTCACTGGGTATTCCTCTGATGATATCAACAAGTTTCTTTGGATGTCGCGA 778

DB	719	TTGAGGCTCCTACTGGTATTCTCTGATGATACCAAGTTTCTTTGGATGGTCCGGA	778
OV	502	lqdlvclvsgvtrhbenrthdclrvveutaelelveclvnenlenrtrmrvtrhtrbrvcmrhc	522

QY 602 leGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrThrProThrG 622

602 TGGGGGGGACCAAGATACAGGCAACATATCAGGAGAACTATTATATCTCCAACTG 838
Db

db 779 TTGGAGGGAGCACAGATACAGGCAACATATCAGGAGATGACTATTATCTCCAACTG 838

Figure 1. The effect of the number of trials on the number of correct responses. The number of correct responses (Y-axis) is plotted against the number of trials (X-axis). The data shows a positive correlation between the number of trials and the number of correct responses, with a linear regression line fitted to the data.

100


```

Db      489 TTATACTCAACTGGGAGTTCGGTGGACCGTGAAGGTTCTCCAGTCTGCTCAACTG 548
Qy      637 sLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgPr 657
Db      549 CCTATGTACAGAAGTGTACTATCGCTTGGACAGGTTTACACAGAAGCCAGCGTCC 608
Qy      657 oProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspVa 677
Db      609 TCCAGGCTTTGACCGTGTCCGAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGAICT 668
Qy      677 lLeuGluGlu 680
Db      669 CTTGGAGGAA 678

RESULT 4
AY418284
LOCUS      2118 bp DNA linear GSS 17-DEC-2003
DEFINITION Homo sapiens ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY418284
VERSION AY418284.1 GI:39774244
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Ranenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2118)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Ranenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source location/Qualifiers
1..2118
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2118
/gene="ITM1"
/locus_tag="HCM650c"

gene

ORIGIN
Alignment Scores:
Pred. No.: 5,08e-202 Length: 2118
Score: 198.00 Matches: 592
Percent Similarity: 96.73% Conservative: 0
Best Local Similarity: 96.73% Mismatches: 10
Query Match: 28.09% Indels: 20
DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x AY418284 (1-2118)

Qy      51 TyrPheAsnTyrArgThrThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsn 70
Db      151 TACTTTTAATATCGGACTACAGGTTCTCGCTGAGGAGGGGTTTATAAATTCATTAAC 210
Qy      71 TrpPheAspArgAlaIatTyrProLeuGlyArgIleGlyGlyThrIleTyrPro 90
Db      211 TGGTTTGTATGACCGAGCGCTGGTACCTTTGGGACGAATCATTTGGAGGAACAATTTACCCA 270

```

```

Qy      91 GlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIle 110
Db      271 GGTTTAATGATCACCTCTGCTGCAATCTACCATGTACTCCATTTTCCACATCCACATC 330
Qy      111 AspIleArgAsnValCysValPheLeuAlaProLeuPheSerSerPheThrSer-IleVa 130
Db      331 GACATTGGGAATGTGTGTGTCTCGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 389
Qy      130 lThrTyrLeu-LeuThrLysGluLeuLysAspAlaGlyValaGlyLeuLeuAlaAlaAla 150
Db      390 CAGGTACCA-CCTTACCAAGAGCTCAAGGATGCAGGGGCTGGGCTTCTTGTCTGCCA 448
Qy      150 etIleAlaValValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 170
Db      449 TGATTGCTGTAGTCTCGATATATCTCCGATCTGTGGCTGGCTCTGTATGATAAAG 508
Qy      170 lYleAlaIlePheCysMetLeuLeuThrTyrTyrMetTyrIleValAlaValLysThrG 190
Db      509 GGATTGCCATCTTTGGCATGCTACTCCTACTACATGTGATCAAGCGAGTAAGACTG 568
Qy      190 lYserIleCysTrpAlaAlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTrpG 210
Db      569 GTTCCATCTGTTGGGAGCTAAGTGTGCCCTTGTCTTATTTCTACATGGTCTCGTATGG 628
Qy      210 lYgLYTyrValPheLeuIleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrG 230
Db      629 GAGGTTATGTGTCTCTGATCAACITTAATCTCTCCAGCTCTCTGCTGATGCTCACAG 688
Qy      230 lYargPheSerHisArgIleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleL 250
Db      689 GCGGTTTCTCACCAGGATCTATGTGGCTACTGTACTGTCTTACTGCTGGCAGCTATAC 748
Qy      250 euSerArg-GlnIleSerPheValGlyPheGlnProValLeuSerSerGluHisMetAla 269
Db      749 TTTCTAT-CCAGATCTCTTGTGGTTCAGCGCTGCTCTTCATCAGAGCAATGGCA 807
Qy      270 Gly-PheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrIleArgSerly 289
Db      808 GC-CTTTGGGGTCTTTGTGCTCTGCCAGATCCATGCCCTTTGTGGATTACCTGGCAGCAA 866
Qy      289 sleuAsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPheVa 309
Db      867 GTTGAATCCACAACTTTGAAGTTCTTTCCGGAGGCTCATCTCTCTGTAGGCTTTGT 926
Qy      309 lleuLeuThrValGlyValAlaLeuLeuMetLeuThrGly----LysIleSerProThrThr 328
Db      927 CCTTCTCACCGTGGGAGCTCTCTCNN-NN-NN-NGAAAAAATATCTCCCTGGACGG 982
Qy      328 lYargPheTyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleIleAlas 348
Db      983 GGGCTTTCTACTCGTGTGGATCCCTCTTATGTAGAACAAACATCCCATCATTTGCTT 1042
Qy      348 erValSerGluHisGlnProThrThrTrpSerSerTyrTyrPheAspLeuGlnLeuLeuV 368
Db      1043 CTGTGTCTGAGCATCAGCCCAACACCTGGTCTCTATCTATTTTGACCTCAGCTCTCTCG 1102
Qy      368 alPheMetPheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIleP 388
Db      1103 TCTTCATGTTTCCAGTGGCTCTCTATTACTGCTTTAGCAACCTGTCTGTATGCCGGATT 1162
Qy      388 heIleIleMetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetL 408
Db      1163 TTATCATCATGTATGTGTGACCAAGCTGTACTTTTTCAGCTGTATGTGTGTGTGTGTAT 1222
Qy      408 euValLeuAlaProValMetSer-IleLeuSerGlyIleGlyValSerGlnValLeuSer 427
Db      1223 TAGTGTGGCACTGTATGTG-CATCTCTCTGCGATTGGAGTCTCCAGGTCGTCTCC 1281
Qy      428 ThrTyrMetLysAsnLeuAspIleSerArgProAspLysLysSerLysLysGlnGlnAsp 447
Db      1282 ACATACATGAAGATCTGGACATAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1341

```

QY 448 SerThrTyrProIleLysIle-GluValAlaSerGlyMetIleLeuValMetAlaPheBp 467
 Db 1342 TCCACCTACCTATTAAAG-TGAAGTGGCAAGTGGGATGATATCTGGTCATGCTTCTT 1400
 QY 467 eLeuIleThrTyrPheHisSerThrTyrPValThrSerGluAlaTyrSerSerProSe 487
 Db 1401 TCTCATCACCTACACCTTTCATCAACCTGGGTGACGAGGAGGCTTCTTCTCGTC 1460
 QY 487 rIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgI 507
 Db 1461 CATGTGATATCTCCCGGTGGGTGGGATGGCAGTAGGATCATATTTGATGACTCCGAGA 1520
 QY 507 uAlaTyrTyrTyrLeuArgHisAsnThrProGluAspAlaLysValMetSerTyrPAs 527
 Db 1521 AGCATATTATTGGCTTCGTCTATAATACTCCAGAGGATGCGAAGTCTATGCTCGTGGA 1580
 QY 527 pTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTr 547
 Db 1581 TTATGGCTATCAGATTACAGTATGGCAACCCGAAACATTTTAGTGGAATAAACACATG 1640
 QY 547 pAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGlyLysAlaTy 567
 Db 1641 GAATAATACCATATTTCTCGATGAGGAGGCAATGGCGTCCACAGAGGAAAGCCTA 1700
 QY 567 rGluIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyLeuThrGI 587
 Db 1701 TGAGATCATGAGGAGCTCGATGTCAGTATGCTGGTCAATTTTGGAGGCGCTACTGG 1760
 QY 587 yTyrSerSerAspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThrAs 607
 Db 1761 GTATTCCTCTCATGATATCAACAGTTTCTTGGATGTCGCGATTGGAGGAGCACACA 1820
 QY 607 pThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThrGlyGluPheArgValAs 627
 Db 1821 TACAGGCAACATATCAAGGAGATGACTATTATCTCAACCTGGGAGTTCGGTGTGA 1880
 QY 627 pArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPh 647
 Db 1881 CCGTAGAGTTCCTCAGTGTGCTCACTGCTCATGCTCAAGATGTGTACTATCGCTT 1940
 QY 647 eGlyGlnValTyrThr 652
 Db 1941 TGGACAGGTTTACACA 1956

RESULT 5
 BI859687
 LOCUS 60388275F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5394834 5',
 DEFINITION mRNA sequence.
 ACCESSION BI859687
 VERSION BI859687.1 GI:16000434
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 770)
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LHAM12006 Row: f Column: 19
 High quality sequence stop: 620.
 Location/Qualifiers
 1..770

FEATURES
 source

LOCUS AU123167
 DEFINITION AU123167 NT2RM1 Homo sapiens cDNA clone NT2RM1000838 5', mRNA
 sequence.

LOCUS AU123167 686 bp mRNA linear EST 01-AUG-2002

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5394834"
 /tissue_type="mammary adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_87"
 /notes="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dr primed.
 Average insert size 1.383 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores: 3.02e-186 Length: 770
 Pred. No.: 183.00 Matches: 209
 Score: 183.00 Conservative: 0
 Percent Similarity: 99.05% Mismatches: 0
 Best Local Similarity: 99.05% Indels: 2
 Query Match: 25.96% Gaps: 0
 DB: 4
 US-10-028-384-12 (1-705) x BI859687 (1-770)
 QY 462 LeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTyrPValThrSerGlu 481
 Db 2 CTGTCATGCTTCTTCTTCATCACCCTACACCTTTCATTCACCTGGGTGACCACTGAG 61
 QY 482 AlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIle 501
 Db 62 GCCTACTCTTCTCCGTCCTTACTATCTGCCCGTGGTGGGATGGCAGTAGGATCATTA 121
 QY 502 PheAspAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProGluAspAlaLys 521
 Db 122 TTTGATCACTTCCGAGAGCATATTATGGCTTCGTCATAAATCTCCAGAGTAGCGAAG 181
 QY 522 ValMetSerTyrPheTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeu 541
 Db 182 GTCATGTCCTGGTGGGATTATGCTATCAGATTACAGCTATGGCAACCGAACAAATTTA 241
 QY 542 ValAspAsnAsnThrTyrPheAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSer 561
 Db 242 GTGGACAATAACACATGGAATAATATCCCATATTTCTCGAGTAGGGCAGCAATGGCGTCC 301
 QY 562 ThrGluGlyLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIle 581
 Db 302 ACAGAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGTGCTGTCATT 361
 QY 582 PheGlyGlyLeuThrGlyTyrSerSerAspAspIle-AsnLysPheLeuTyrMetValAr 601
 Db 362 TTTGGAGGCTCACTGGGTATTCCTCTGATGATATCAAAACAAGTTTCTTTGGATGGTCGG 421
 QY 601 gIleGlyLysThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProTh 621
 Db 422 GATTGGAGGAGCACAGATACAGCAACATATCAAGGAGATGACTATTATCTCCAC 481
 QY 621 rGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrIly 641
 Db 482 TGGGAGTTCCGCTGGACCGTGAAGGTTCTCCAGTGTGCTCAACTGCTCATGTACAA 541
 QY 641 sMetCysTyrTyrArgPheGlyGlnValTyrGlu-AlaLysArgProGlyPheA 661
 Db 542 GATGTGTACTATCGCTTTGGACAGTTTACACAGAAAGCGGCTCTCCAGGCTTGG 601
 QY 661 spArgValArgAsnAlaGluIleGlyAsn 670
 Db 602 ACCGTGTCCGAAATGTGAGATCGGAAT 630

RESULT 6

LOCUS AU123167
 DEFINITION AU123167 NT2RM1 Homo sapiens cDNA clone NT2RM1000838 5', mRNA
 sequence.

LOCUS AU123167 686 bp mRNA linear EST 01-AUG-2002

```

ACCESSION      AUI23167
VERSION        AUI23167.1  GI:10947883
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens

REFERENCE
AUTHORS      Ota,T., Nishikawa,T., Suzuki,Y., Kawai,Y., Ishii,S., Saito,K.,
              Nakamura,I., Nagai,T., Sugano,S. and Isogai,T.
TITLE        HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Kawai,Y.,
              Ishii,S., Saito,K., Nakamura,I., Nagai,T., Sugano,S., Isogai,T.)
JOURNAL      Unpublished (2000)
COMMENT      Contact: Takao Isogai
              Genomics Laboratory
              Helix Research Institute
              1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
              Tel: 81-438-52-3975
              Fax: 81-438-52-3986
              Email: genomics@hri.co.jp
              HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
              Research Institute; cDNA library construction; Department of
              Virology, Institute of Medical Science, University of Tokyo, and
              Helix Research Institute.
FEATURES
source
1..686
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="NT2RM1000838"
   /cell_type="teratocarcinoma"
   /cell_line="NT2"
   /clone_lib="NT2RM1"
   /notes="Vector: pUC19FL3; mRNA from uninduced NT2 neuronal
   precursor cells"

ORIGIN
Alignment Scores:
Pred. No.:      5,65e-182      Length:      686
Score:          179.00      Matches:      179
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    25.39%      Indels:      0
DB:              1      Gaps:      0

US-10-028-384-12 (1-705) x AUI23167 (1-686)
QY 502 PheAspAspPheArgGluAlaValTyrTyrTyrLeuArgHisAsnThrProGluAspAlaLys 521
Db 3 TTTGATGACTTCGAGAGACATATTAATGCTTCGTATATACATCCAGAGATCGGAG 62
QY 522 ValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeu 541
Db 63 GTCATGTCTCCGTGGGATTATGCTATACAGATTACAGCTATGCGAAACCGAACAAATTTA 122
QY 542 ValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSer 561
Db 123 GTGGACATTAACATGGAATATATACCATATTTCTCGAGTAGGCGAGCATGGCGTCC 182
QY 562 ThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValIleuValIle 581
Db 183 ACAGAGAAAGAACCTTATGAGATCATGAGGAGCTCGATGTGATGATGCTGCTGCTATT 242
QY 582 PheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArg 601
Db 243 TTTGGAGCCCTCACTGGGTATCTCTGATGATATACACAGCTTCTTTGGATGGTCCGG 302
QY 602 IleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThr 621
Db 303 ATTGGAGGAGCAGATACAGCAACATATCAAGGAGATGACTATATTACTCCAACT 362
QY 622 GlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLys 641

```

```

363 GGGGAGTTCCGTGTGGACCGCTGAAGGTTCTCCAGTGTGCTCACTGCCTCATGTACAAAG 422
QY 642 MetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAsp 661
Db 423 ATGTGTTACTATCGGTTTGGACAGGTTTACACAGAACCAAGCGTCTCCAGGCTTTGAC 482
QY 662 ArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
Db 483 CGTGTCCCAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGTATGCTCTGGAGGAA 539

RESULT 7
CN265339
LOCUS      CN265339
DEFINITION 1700059935691 GRN_PRENEU Homo sapiens cDNA 5', mRNA sequence.
ACCESSION  CN265339
VERSION    CN265339.1  GI:47281753
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens

REFERENCE  1 (bases 1 to 541)
AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
           Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
           Lebkowski,J and Stanton,L.W
TITLE     Transcriptome characterization elucidates signaling networks that
           control human ES cell growth and differentiation
JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT   Contact: Brandenberger R
           Regenerative Medicine
           Genon Corporation
           230 Constitution Drive, Menlo Park, CA 94025, USA
           Tel: 650 473 8658
           Fax: 650 473 7760
           Email: rbrandenberger@genon.com
           Insert Length: 541 Std Error: 0.00.
           Location/Qualifiers
             1..541
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /tissue_type="embryonic stem cell, retinoic acid and
               mitogen-treated hES cell line H7"
               /clone_lib="GRN_PRENEU"
               /note="cligo dt primed, full-length enriched cDNA library
               from hES cell line H7 (p29) maintained in feeder-free
               conditions. Embryoid bodies were generated in the presence
               of all-trans retinoic acid and mitogens."

ORIGIN
Alignment Scores:
Pred. No.:      5,39e-181      Length:      541
Score:          178.00      Matches:      178
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    25.25%      Indels:      0
DB:              7      Gaps:      0

US-10-028-384-12 (1-705) x CN265339 (1-541)
QY 455 GluValAlaSerGlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHis 474
Db 6 GAAGTGGCAAGTGGGATGATGACTGTCATGCTGCTTTCTTCTCATCACCTACACCTTCAT 65
QY 475 SerThrTrpValThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGly 494
Db 66 TCACCTGGGTGACAGTGGAGGCTACTCTTCCTCCGTCCTTACTACTCTGCGCGTGT 125
QY 495 GlyAspGlySerArgIleIlePheAspPheArgGluAlaTyrTyrTrpLeuArgHis 514
Db 126 GGGGATGGCAGTAGGATCATATTGATGACTTCGAGAGCATATTATTGGCTTCGTGAT 185
QY 515 AsnThrProGluAspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAla 534

```

```

Db 186 AATACTCCAGAGGATCGAAGGTCATGTCCTGGTGGGATTATGCTATCAGATTACAGCT 245
Qy 535 MetAlaAsnArgThrIleValAspAsnSerThrTrpAsnSerThrIleValSerArg 554
Db 246 ATGGCAACCGACCAATTTTATGACACATACATGGAATATATCCCATATTTCTCGA 305
Qy 555 ValGlyGlnAlaMetAlaSerThrGluGluLysAlaValTrpGluMetArgGluLeuAsp 574
Db 306 GTAGGCGCAGGAATGGCGTCACAGAGGAAAAAGCCTATGAGATCATGAGGAGGCTCGAT 365
Qy 575 ValSerTrpValLeuValIlePheGlyLeuThrGlyTrpSerSerAspIleLeuAsn 594
Db 366 GTACACTATGCTGGTCATTTTGGAGGCTCTACTGGTATTCCTCTGATGATATCAAC 425
Qy 595 LysPheLeuTrpMetValArgIleGlySerThrAspThrGlyLysHisIleLysGlu 614
Db 426 AAGTTTCTTGGATGCTCGGATGGAGGAGGACAGATACAGGCAACATATCAAGGAG 485
Qy 615 AsnAspTyrTrpProThrGlyGluPheArgValAspArgGluGlySerPro 632
Db 486 AATGACTATTTATCTCCACATGGGAGTTCCGTGGACCGTGAGGTTCTCCA 539

RESULT 8
AI905439
LOCUS RC-BT091-200199-026 BT091 Homo sapiens cDNA, linear EST 30-MAR-2000
ACCESSION AI905439
VERSION AI905439.1 GI:6495826
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 692)
Dias Neto,E., Garcia Correa,R., Vertovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., de Oliveira,P.S., Bucher,P., Jorgensen,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/seq/gethtml.pl?t1=RC&t2=RC-BT091-026.html
&t3=200199&t4=1)
Seq primer: puc 18 forward.
Location/Qualifiers
1..692
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="female"
/dev_stages="Adult"
/clone_lib="BT091"
/note="Organ: breast; Vector: puc18; Site: 1; SmaI; Site 2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No.
196,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low

```

stringency conditions."

```

ORIGIN
Alignment Scores:
Pred. No.: 6.95e-181 Length: 692
Score: 178.00 Matches: 178
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.25% Indels: 0
DB: 1 Gaps: 0

US-10-028-384-12 (1-705) x AI905439 (1-692)
Qy 476 ThrTrpValThrSerGluAlaValTrpSerSerProSerIleValLeuSerAlaArgGlyGly 495
Db 20 ACCCTGGGTGACCATGAGGCTACTCTCTCCGTCATTGTACTATCTCCCGTGGTGGG 79
Qy 496 AspGlySerArgIleIlePheAspAspPheArgGluAlaValTrpTrpLeuArgHisAsn 515
Db 80 GATCGCAGTAGGATCATATTTGATGACTTCCGAGAGCATATTTATGGCTTCGTATAT 139
Qy 516 ThrProGluAspAlaLysValMetSerTrpTrpAspTyrGlyTrpGlnIleThrAlaMet 535
Db 140 ACTCCAGAGGATCGAAGGTCATGCTCTGCTGGGATTATGGCTATCAGATTACAGCATG 199
Qy 536 AlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgVal 555
Db 200 GCAACCCGACCAATTTTAGTGACATACACATGGAATATATACCATATTTCTCGAGTA 259
Qy 556 GlyGlnAlaMetAlaSerThrGluGluLysAlaValTrpGluLeuMetArgGluLeuAspVal 575
Db 260 GGGCAGGCAATGGCGTCCACAGAGGAAAAAGCCTATGAGATCATGAGGAGGCTCGATGTC 319
Qy 576 SerTrpValLeuValIlePheGlyGlyLeuThrGlyTrpSerSerAspAspIleAsnLys 595
Db 320 AGCTATGCTGGTCATTTTGGAGGCCCTCAGTGGTATTCCTCTCATCATATCAACAG 379
Qy 596 PheLeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsn 615
Db 380 TTTCTTTGGATGGTTCGGATGGAGGAGCAGACATACAGGCAACATATCAAGGAGAA 439
Qy 616 AspTrpTrpThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeu 635
Db 440 GACTATTTATCTCCACTCGGAGTTCCTGGTGGACCGTGAAGGTTCTCCAGTGTGCTC 499
Qy 636 AsnCysLeuMetTrpLysMetCysTrpTrpArgPheGlyGlnValTrpThrGlu 653
Db 500 AACTGCCTCATGTACAGATGTCTACTATCGCTTTGGACAGGTTTACACAGAA 553

```

CA487523 994 bp mRNA linear EST 14-NOV-2002
AGENCOURT_10809261 MAPcL Homo sapiens cDNA clone IMAGE:6719013 5',
mRNA sequence.

CA487523
CA487523.1 GI:24947110

EST.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 994)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

JOURNAL

COMMENT

Email: cgabs@xmail.nih.gov

Tissue Procurement: Kristi A. Eglund, Ira Pastan

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LHM14276 row: d column: 21
High quality sequence stop: 623
Location/Qualifiers

FEATURES

source

1..994
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6719013"
/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
Hs578T-HME1, LNCaP"
/lab_host="EMDH10B"
/lab_lib="NAPCH"
/note="vector: pCMV-SPORT6; Site_1: EcoRV; Site_2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkook Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:
Pred. No.: 1,296-165 Length: 994
Score: 164.00 Matches: 164
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.26% Indels: 0
DB: 6 Gaps: 0

US-10-028-384-12 (1-705) x CA487523 (1-994)

QY 517 ProGluAspAlaLysValMetSerTrpTyrPheGlyTyrGlnIleThrAlaMetAla 536
DB 1 CCAGAGATCGGAGGTCATGCTCTGGTGGATTATGGTATACAGTATGCA 60
QY 537 AsnArgThrIleValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGly 556
DB 61 AACCGAACATATTTAGTGGCAATAACACATGGATAATACCATTATTCGAGTAGGG 120
QY 557 GlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSer 576
DB 121 CAGGCATGGCTCCACAGAGGAAAGGCTATGATCATGAGGAGCTCGATGTCAGC 180
QY 577 TyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPhe 596
DB 181 TATGTCTGTGTCATTTTGGAGGCTCCTGCTGATGATATCAACAAGTTT 240
QY 597 LeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAsp 616
DB 241 CTTTGGATGTCGGATTTGAGGAGGACACAGATACAGGCAACATATCAAGGAGATGAC 300
QY 617 TyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsn 636
DB 301 TATTATCTCAACTGGGGAGTTCCGTGTGGACCGTGAAGGTTCCAGTGTGCTCAAC 360
QY 637 CysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArg 656
DB 361 TGCCTCANGTACAGATGTGTTACTATCGCTTTGGACAGGTTTACACAGAGCCCAAGGTT 420
QY 657 ProProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAsp 676
DB 421 CCTCCAGGCTTTCACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGT 480
QY 677 ValLeuGluGlu 680
DB 481 GTCCTGGAGGA 492

RESULT 10

BX284031

LOCUS

DEFINITION BX284031 NCI_CGAP_skn4 Homo sapiens cDNA clone IMAGE998E0410606 ;

484 bp

mRNA

linear

EST 05-MAR-2003

IMAGE:4750827, mRNA sequence.

BX284031

VERSION BX284031.1 GI:28848485

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 484)
Radelof, U., Schneider, D. and Korn, B.
Human Unigeneset - RZPD3
Unpublished (2003)
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGE998E0410606.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-
bin/showLib.pl.cgi?response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
Contact RZPD (clone@rzpd.de) for further information. Seq primer:
SP6, Primer sequence: ATTAGTGACACTATAG.

FEATURES

source

1..484
/location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE998E0410606 ; IMAGE:4750827"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (TI phage-resistant)"
/clone_lib="NCI_CGAP_skn4"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dr.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Alignment Scores:
Pred. No.: 1,116-162 Length: 484
Score: 161.00 Matches: 161
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 22.84% Indels: 0
DB: 5 Gaps: 0

US-10-028-384-12 (1-705) x BX284031 (1-484)

QY 455 GluValAlaSerGlyMetIleLeuValMetAlaPhePheLeuIleThrTyrPheHis 474
DB 1 GAAGTGGCAAGTGGGATGATGATCTGGTCATGGCTTCTTCTCATCCTACCTACCTTCAT 60
QY 475 SerThrTrpValThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGly 494
DB 61 TCACCTGGGTGGACCAAGTGGAGGCTCTCTCTCCGTCATTGTACTATCTGCCGTGGT 120
QY 495 GlyAspGlySerArgIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHis 514
DB 121 GGGGATGGCAGTAGGATCATATTTGATGACTTCCGAGACATATTTGGCTTCGTCTAT 180
QY 515 AsnThrProGluAspAlaLysValMetSerTrpTyrAspTyrGlyTyrGlnIleThrAla 534
DB 181 ATACTCCAGAGGATGCCAAGGTCTATGCTCTGGTGGGATTTATGGCTATCAGATTACAGCT 240
QY 535 MetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArg 554
DB 241 ATGGCAACCCGAAACAAATTTTATGTGGACAATAACACATGGGAATAATACCCCATATTTCTCGA 300

QY 555 ValGlyGlnAlaMetAlaSerThrCluGluLysAlaThrGluLeuMetArgGluLeuAsp 574
 Db 301 GTAGGCGAGGAGTGGCGTCCACAGAGGAAAGCCATGAGATCATGAGGGAGCTCGAT 360
 QY 575 ValSerTyrValLeuValPheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsn 594
 Db 361 GTACAGTATGTGCTGGTCAATTTTGGAGGCGCTCACTGGTATTCCTCTGATGATATCAAC 420
 QY 595 LysPheLeuTyrMetValArgGlyGlySerThrAspThrGlyLysHisIleLysGlu 614
 Db 421 AAGTTTCTTTGGATGGTCCGATTTGGAGGAGGACAGATACAGCAACATATCAAGGAG 480
 QY 615 Asn 615
 Db 481 AAT 483
 RESULT 11
 LOCUS CK838051 733 bp mRNA linear EST 04-MAR-2004
 DEFINITION 4083016 BARC 8BOV Bos taurus cDNA clone 8BOV_10J04 5', mRNA
 ACCESSION CK838051
 VERSION CK838051
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.
 1 (bases 1 to 733)
 Baumann,R.G., Baldwin,R.L., Sonstegard,T.S., Van Tassell,C.P. and
 Matukumalli,L.K.
 Construction and Analysis of a cDNA Library Generated From
 Intestinal Muscle and Epithelial Tissues of Holstein Cattle
 Unpublished (2004)
 CONTACT: Richard G. Baumann
 Bovine Functional Genomics Lab
 ANRI
 BLDG 162: BARC-EAST, Beltsville, MD 20705, USA
 Tel: 3015048604
 Fax: 3015048744
 Email: rbaumann@anri.barc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred
 0.000925 using options -trim.alt -trim.fasta. Vector identified
 by cross match using options -minmatch 12 -minscore 12
 Plate: 10 row: J column: 04
 Seg primer: CTTATTAGGTGACACTATAGAAC
 High quality sequence stop: 733.
 Location/Qualifiers
 1..733
 /organism="Bos taurus"
 /mol_type="mRNA"
 /strain="Holstein"
 /db_xref="taxon:9913"
 /clone="8BOV_10J04"
 /sex="Female"
 /tissue_type="Epithelial, Muscle"
 /dev_stage="Lactating, Neonatal"
 /lab_host="DH10B TONa"
 /clone_lib="BARC 8BOV"
 /note="Organ: Intestine; Vector: pCMVSPORT6.1; Site: 1:
 NotI; Site 2: EcoRI; Normalized cow cDNA intestinal
 library in pCMVSPORT6.1, constructed from equimolar mRNA
 pools derived from 5 sources, 4 lactating intestinal, 1
 neonatal intestinal 4/5 Lactating, proximal Duodenum,
 Jejunum, Distal ileum, Colon, 1/5 Neonatal, Proximal
 Duodenum, Jejunum, Distal ileum"
 ORIGIN
 Alignment Scores:
 Pred. No.: 733
 Score: 160.00
 Length: 160
 Matches: 160

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 22.70% Indels: 0
 DB: Gaps: 0
 US-10-028-384-12 (1-705) x CK838051 (1-733)
 QY 455 GluValAlaSerGlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHis 474
 Db 203 GAAGTGGCAAGTGGCATGATGCTGGTCAATGGCTTTCTTCATTAACCTACACCTCCAC 262
 QY 475 SerThrTyrValThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGly 494
 Db 263 TCGACCTGGGTGACCAAGTGGAGGCTACTCTCTCTCCCTCCATTGTGCTGCTGCTGTGT 322
 QY 495 GlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHis 514
 Db 323 GGAGATGGCAGTAGGATCATTTTGTGATTTTCGAGAAGCGTACTATTGGCTTCGTCCAC 382
 QY 515 AsnThrProGluAspAlaLysValMetSerTyrTrpAspTyrGlyTyrGlnIleThrAla 534
 Db 383 AATACTCCAGAGGATGCGAAGGTCAATGCTGGGATTATGGCTACCAAGATTACAGCT 442
 QY 535 MetAlaAsnArgThrIleLeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArg 554
 Db 443 ATGGCGAATCGGACGATTTTAGTGGATAATAACACCTGGATAATACCATATATCTCGA 502
 QY 555 ValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluLeuMetArgGluLeuAsp 574
 Db 503 GTAGGGAGGACATGGCATCCACAGAGAAAGAAAGCCATGAGATCATGAGGAGGCTTGAT 562
 QY 575 ValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsn 594
 Db 563 GTACAGTATGCTGGTCAATTTTGGAGGCTCACTGGGTATTTCTTCAGATGACATCAAC 622
 QY 595 LysPheLeuTyrMetValArgIleGlySerThrAspThrGlyLysHisIleLysGlu 614
 Db 623 AAATTTCTGTGGATGTCGGGATTCGGGAGGAGGACAGATACAGGAAACACATCAGGAG 682
 RESULT 12
 LOCUS CA454755 914 bp mRNA linear EST 12-NOV-2002
 DEFINITION AGENCOURT_10763160 MAPcL Homo sapiens cDNA clone IMAGE:6721291 5',
 mRNA sequence.
 ACCESSION CA454755
 VERSION CA454755.1 GI:24904799
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 914)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: Kristi A. Egland, Ira Pastan
 cDNA Library Preparation: Invitrogen Corp
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM14282 row: c column: 19
 High quality sequence stop: 687.
 Location/Qualifiers
 1..914
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6721291"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,"
 FEATURES
 source

hTERT-HMEL, LNCap"
/lab_host="EMD10B"
/clone_lib="NAPCh"
note=vector: pCMV-SPORT6; Site_1: EcoRV; Site_2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kirsti A. Eglund, James J. Vincent, Robert Strausberg,
Kung'ook Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

Alignment Scores:		
Pred. No.:	4,31e-158	914
Score:	157.00	204
Percent Similarity:	99.03%	Conservative: 0
Best Local Similarity:	99.03%	Mismatches: 1
Query Match:	22.27%	Indels: 2
DB:	6	Gaps: 0
US-10-028-384-12 (1-705)	X	CA454755 (1-914)

US-10-028-384-12 (1-705) x CA454755 (1-914)

476 ThrTrpValThrSerGluAlaTyrSerSerProSerIleValLeuSerLysArgGlycyl 495
36 ACCTGGGGACAGTAGGGGCTACTCTTCCTCGCATTTGATATCTCCCGGTGGG 95
436 AspGlySerArgIleIlePheAspAppPheArgGluAlaTyrTrpPheuArgHisAsn 515
96 GATGGCAGTAGGATCATATTGTCATCTCCGAGAAGCATATTATTGGCTGTCATAT 155

QY 516 ThrProGluAspAlaIysValMet-SerTrpTrpAspTyrGlyTyrGlnIleThrAlaIve. 535
 |||||
 Db 156 ACTCCAGAGGATGCGAAGGTCCTT-GTCCCTGGTGGGATTATGGCTATCATGATTACAGCTAT 214

	QY	535	TAAAGATGTTTLLLEUVAASPDASNTHIRIPASNASHNHRLISERATGVA	555
Dd		215	GGCAACCGAACAAITTTTAGTGACATAAACAATGGATAATACCCTATTCTCGAGT	274

555 1G1yGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspVa 575
275 AGGCAGGCGAATGGCGTCCACAGAGGAGAAAAGCCTATGAGATCACTAGGGGCTCGATGT 314

Qy 575 lSerrTyrralleuvallephGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnIly 595

QY 595 sPheLeuTrpMetValArgIleGlyGlySerThrAsp^{trh}GlyLysHisIleLysGluAs 615

Qy 615 nAspTyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeu 635

	435	IGACIATIALAACICCAACAGCGGGAAATCCGCTGTGGACCCGAGGAAGTTCCTCAGAGCTGCCT	319
OY	635	uANcysLeuMetTyrLysMetCystTyrTyxArgpheGlycInValTyrThrGlutAlaly	655

515	CAACTGCTCATGTACAAGATGTGTTACTATCGCTTTTGGACAGGTTTACACAGAAGCCAA	571
DB		
QY	655	sATcPrpPrpCgLVpHeAspATcValATcAsnAlaIaGluIleGLVAsnLVsAspPheGluIuLe
	675	

Db 575 GCCTCTCCAGGCTTGACCGTGCAGAAATCTGAGATTGGGAATAAGACTTTGAGCT 634

Qv 575 uRenVallouClnClu 580

Db
635 TGATGCTCTGAGGAA 650

LOCUS	AI437149	AI437149	515 bp	mRNA	linear	EST 07-JUN-2001
DEFINITION	fb38c11.y1 Zebrafish Washu NP1MG EST Danio rerio cDNA clone IMAGE:3714164 5', similar to SW:5773 MOUSE P46978 OLIGOSACCHARYL TRANSFERASE S77, similar to SW:101012 HOMOLOG, mRNA sequence					

STUDY OF THE EFFECTS OF THE

ACCESSION A1437149
 VERSION A1437149.1 GI:4288291
 KEYWORDS EST.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 615)
 REFERENCE Clark,M.,Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,I., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson.R.
 Washu Zebrafish EST Project 1998
 UNPUBLISHED (1998)
 CONTACT: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrfish@watson.wustl.edu
 CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and RessourcenzentrumPrimardatenbank, Berlin, Germany (web address: www.rzpd.de)
 POSSIBLE REVERSED CLONE: similarity on wrong strand
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 497

```

FEATURES
source
FOLDING=NO..
Location/Qualifiers
1. .615
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:3714164"
/sex="mixed"
/tissue type="26 somite embryos, adult livers, shield stage embryos"
/lab host="Xl1-blue MRF"
/clone_lib="zebrafish Washu MPIMG EST"
/note="vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer [5'-pgactgattctagatcgacggcgcgtttttttt]; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Juehrach lab; ICRR, London and Max Planck Institut fuer Molekulare Genetik,Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control"
```

ORIGIN	Alignment Scores:	4.24e-156	Length:	615
	Pred. No.:	153.00	Matches:	155
	Score:	100.00%	Conservative:	0
	Percent Similarity:	100.00%	Mismatches:	0
	Best Local Similarity:	21.99%	Indels:	0
	Query Match:		Gaps:	0
	Ds:	1		

adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."

TITLE A Comprehensive Collection of Chicken cDNAs

JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392

COMMENT

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 0161208930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source

1..743
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layard"
/db_xref="taxon:9031"
/clone="CHES7677n22"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEORBN21"
/note="Organ: ovary; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

ORIGIN

Alignment Scores:
Pred. No.: 5.1e-156 Length: 743
Score: 155.00 Matches: 228
Percent Similarity: 98.28% Conservative: 0
Best Local Similarity: 98.28% Mismatches: 2
Query Match: 21.99% Indels: 4
DB: 5 Gaps: 0

US-10-028-384-12 (1-705) x BU471816 (1-743)

Qy 380 SerAsnLeuSerAspAlaArgllePheIleMetTyrGlyValThrSerMetTyrPhe 399
Db 5 AGCAATCTCTCAGATGCCGCAATTTTATCATCATGTATGGCGTACCTCATCTACTTTC 64
Qy 400 SerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG1 419
Db 65 TCTGCTGTAATGCTGCGGTGATGCTGGTGGTGGCCCGGTGATGTG-CATCTTTTCGG 123
Qy 419 YileGlyValSerGlnValLeuSerThrTyrMetTyrAsnLeuAspIleSerArgProAs 439
Db 124 CATCGCGGTTCCTCAGGTGTGTTCACCTACATGAAGAACCTGGACATCAGCGGCCGAGA 183
Qy 439 pLysLysSerLysLysGlnGluAspSerThrTyrProIle-LysIleGluValAlaSerG 459
Db 184 CAAGAGAGCAAAAGACAGACAGACTCCACCTACCCATCAAAAT-GAAGTTCGCCAGG 242
Qy 459 llyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValT 479
Db 243 GCATGATCCCTGGGTGGTGTTCCTCATCAGTACACCTTCACCTCGACCTGGGTGA 302
Qy 479 hrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerA 499
Db 303 CCAGCGAGGCATACTCATCGCCCTCCATCGTCTGTGAGCGCGCGGGGACGCGCAGCA 362

Qy 499 rglIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluA 519
Db 363 GAATCATCTTTGATGACTTCAGGGAGGCGCTACTACTGCTGGGCACAAACACACCGAGG 422
Qy 519 spAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgT 539
Db 423 AGCCCAAGGTGATGTCCTGCTGGGACTACGGCTACAGATCACCAGCATCGCCATGGCCACCGCA 482
Qy 539 hrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaM 559
Db 483 CCATCTGCTGGTGACAAACACACCTGGGAACACACACGACATCTCCCGTGTGGCAGCGCA 542
Qy 559 etAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValL 579
Db 543 TGGCATCGACGGAGGAGAAAGCTTACGAGATCATGAGAGAGCTGGACGTGAGCTACGTGC 602
Qy 579 euValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpM 599
Db 603 TGGTGTATCTTCGGCGGCTCACCAGCTATTGCTCCGACGACATCAACAAAGTTCCTGTGA 662
Qy 599 etValArgIleGlyGlySerThrAspThrGly 609
Db 663 TGGTGGGCAATCGGCGGCGACGACACAGACGCGGC 694

Search completed: December 15, 2004, 06:14:40

Job time : 5478.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:51:45 ; Search time 7810.27 Seconds
(without alignments)
4268.640 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 3697

Sequence: 1 MKKGFRLSLVEKQDTLLKL.....HMLVRIYKVKDLDNRLSRT 705

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131519_8017/app.query.fasta_1.3740
-DB=GenEmbl -QWTF=fastcap -SUFFIX=rge -MINMATCH=0 -LOOPEXT=0
-UNIQS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIS7=45
-LOCALIGN=200 -THR_SCORES=500 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HAPSISIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn_1_1_20262@runat_14122004_131519_8017 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DGPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3697	100.0	2472	6	AX799092 Sequence
2	3697	100.0	2472	9	L38961 Human putat
3	3657	98.9	2118	9	BT007100 Homo sapi
4	3657	98.9	2118	12	BT008132 Synthetic

RESULT 1
AX799092
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
ORIGIN
Alignment Scores:
Pred. No.:
Score:

AX799092
Sequence 11 from Patent WO03054008.
AX799092
AX799092.1 GI:37605063
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1
Perreault, C. and McBride, K.
Mammalian S1MP protein, gene sequence and uses thereof in cancer therapy
Patent: WO 03054008-A 11 03-JUL-2003;
Compatisgene Inc. (CA)
Location/Qualifiers
1. 2472
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/note="ITM1, accession No NM-002219"

2450 9 BC048348
2458 6 BC034714
2516 6 BC020965
2472 6 CQ27867
2293 9 HUMTMC
2898 9 HSM809254
2435 6 CQ834710
2736 10 BC037612
3094 6 AX799092
3094 10 MUXTRPR
2760 6 CQ492624
3666 10 AK129027
2481 5 BC067313
2567 5 BC063234
2608 5 BC046072
2329 6 CQ834708
2855 6 CQ573675
2470 6 CQ834712
2339 6 CQ834706
6153 6 CQ573674
18405 2 AC017914
18405 3 AC104631
283556 3 AE003589
2555 8 BT012881
2481 9 AX074880
2710 6 AX799084
2756 8 AK099138
4236 10 BC052433
2819 8 AK067609
1522 5 BC064436
1421 5 BC043914
2417 3 AF132552
58.9 2417 6 AX799088
58.9 2699 6 CQ589353
56.9 4922 6 CQ589352
56.9 162921 3 AC007853
56.9 181132 3 AC008206
56.9 227219 3 AE003750
56.2 75650 2 AC018145
56.0 2256 6 AX489023

ALIGNMENTS

BC048348 Homo sapi
CQ834714 Sequence
BC020965 Homo sapi
CQ27867 Sequence
L47337 Homo sapien
BX649102 Homo sapi
CQ834710 Sequence
BC037612 Mus muscu
AX799090 Sequence
L34260 Mus musculu
CQ492624 Sequence
AK129027 Mus muscu
BC067313 Xenopus t
BC063234 Danio rer
BC046072 Danio rer
CQ834708 Sequence
CQ573675 Sequence
CQ834712 Sequence
CQ834706 Sequence
CQ573674 Sequence
AC017914 Drosophil
AC104631 Drosophil
AE003589 Drosophil
BT012881 Lycoperal
AX799082 Sequence
AX074880 Homo sapi
AX799084 Sequence
AK099138 Oryza sat
BC052433 Mus muscu
AK067609 Oryza sat
BC064436 Xenopus l
BC043914 Xenopus l
AF132552 Drosophil
AX799088 Sequence
CQ589353 Sequence
CQ589352 Sequence
AC007853 Drosophil
AC008206 Drosophil
AE003750 Drosophil
AC018145 Drosophil
AX489023 Sequence

Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-12 (1-705) x AX799092 (1-2472)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 107 ATGACTAAGTTTGGATTTCGATTGCTCTATGAGACGACGACACACTTTTGAAGCTT 166
 QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgPheAlaValLeuArg 40
 DB 167 CTCATTCTGCATGCGTCTGCTGATTATCCCTTCACCTCGCTCTGTTGCTGCTGAGA 226
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
 DB 227 TTTGAAAGTGTATCCATGAGTTTGATCCGTACTCTTTAATTATCGGACTACACAGGTTCCGT 286
 QY 61 AlaGluGlyPheThrLysPheHisAsnTyrPheAspArgAlaTyrPheProLeu 80
 DB 287 GCTGAGGAGGGGTTTATAAATCCATAACTGTTTGAAGCCGAGCCTGGTACCCCTTTG 346
 QY 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 DB 347 GGACGAATCATTCGAGGACAAATTTACCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 406
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 DB 407 CATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGTCTGTGTGTTCTCGGCC 466
 QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 DB 467 CCTCTCTTCCTCCCTCACCTCCATCGTCACGTACCTACTCTCTTACCAAGAGGCTCAAGGAT 526
 QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
 DB 527 GCAGGGCTGGGCTTCTGTCTGCTGCATGATTGTCTGATGTTCTCGGATATATCTCCGA 586
 QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 DB 587 TCTGTGCTGGCTCCTATGATATGAATGAAGGATTGCCATCTTTTGATGCTACTCACCTAC 646
 QY 181 TyrMetThrLysAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaLeu 200
 DB 647 TACATGGGATCAAGGACGTAAGACTGGTTCCATCTGTGGCAGCTAAGTGTGCCCTT 706
 QY 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuLeuAsnLeuLeuPro 220
 DB 707 GCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTCTTCTGATCAACTTAATTCCT 766
 QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 DB 767 CTCACAGTCTCTGTGCTGATGCTCACAGGCGGTTTCTCTACCGGATCTATGTGCGCTAC 826
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 DB 827 TGTACTGTTTACTGCTCGCTGCTATATCTTCTAGGCAGATCTCCTTGTGGGTTTCCAG 886
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 DB 887 CCTGCTCTCTCATCAGACACATGGCAGGGGTTTGGGGTCTTTGGTCTCTGCGACATCCAT 946
 QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 DB 947 GCCTTTTGTGATTACCTGGCAGCAAGTTCAATCCACAACTTTTGAAGTCTTTTTCGG 1006
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 DB 1007 AGCGTCACTCTCTGGTAGGCTTTGCTTCTTCTACCGTGGGAGCTCTCTCTCATGCTGACA 1066
 QY 321 GlyLysIleSerProThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340

DB 1067 GGAAAAATATCTCCCTGACCGGGCGTTTCTACTCTACTGCTGGATCCCTCTTATGCTAAG 1126
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrThrPheSerSerTyr 360
 DB 1127 AACAAATCCCATCATTCCTGCTGCTGAGCATCAGCCCAACCTGGTCTCTCATAC 1186
 QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 DB 1187 TATTTTGAACCTGCAGCTCCTCTCTCATGTTTCCAGTTGGCTCTTATTAAGTCTTAGC 1246
 QY 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1247 AACCTGTCTGATGCCGAGTTTATCATCATGATGCTGTGACCAAGCATGACTTTTCA 1306
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1307 GCTGTAATGGTGGCTTAATCTAGTGTGGCACCTGTTATGAGCATTTCTCTGTCGATT 1366
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 DB 1367 GGAGTCTCCAGGTGCTGCTCCACATCATGAGAAATCTGGACATTAAGTCCGCCACAG 1426
 QY 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 DB 1427 AAGAGCAAGAAACACAGGATCCACCTACCTTAAAGATTGAAGTGGCAAGTGGGATG 1486
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrThrValThrSer 480
 DB 1487 ATACGTGTCATGGCTTCTTCTCATCTACCTACCTTTCATTCACCTGGGTGACCAT 1546
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 DB 1547 GAGGCTACTTCTCTCCGTCATTTGCTACTATCTGCCGCTGGTGGGATGGCAGTAGATC 1606
 QY 501 IlePheAspAspPheArgGluAlaTyrTyrTyrPheArgHisAsnThrProGluAspAla 520
 DB 1607 ATATTGTGATGCTTCCGAGAAAGCATATTTGGCTTCTGCTCATTAATCTCCAGAGGATGG 1666
 QY 521 LysValMetSerTyrPheTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 DB 1667 AAGGTCTATGCTGCTGGATTTATGGTATCATGATACAGCTATGGCAACCAAT 1726
 QY 541 LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1727 TTATGTGACATATACATATGGAATATATCCCATATTTCTCGAGTAGGAGGCAATGGG 1786
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1787 TCCACAGAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGCTAGCTATGTGCTGGTC 1846
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTyrMetVal 600
 DB 1847 ATTTTGTGAGGCTCTACCTGGGTATTTCTCTGATGATATCAACAGTTTCTTTGGAAGTCT 1906
 QY 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 DB 1907 CGGATTGGAGGAGCACAGATACAGGCAACATATCAAGGAGATGACTATTATATATCCA 1966
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 DB 1967 ACTGGGAGTTCCGTTGGACCTGGAAGTTTCTCCAGTGTGCTCAACTGCCTCATGTATC 2026
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 DB 2027 AAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAACCAAGCGTCTCTCCAGCTTT 2086
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 DB 2087 GACCTGTCCGAATGCTGAGATTGGGAATAGAGACTTTGAGCTTGTATGCTCTGGAGGAA 2146
 QY 681 GlyTyrThrThrGluHisThrPheValArgIleTyrLysValLysAspLeuAspAsnArg 700
 DB 2147 GGCTATACCAAGAACATTTGGCTGCTCAGGATATACAAGGTAAAGGACCTGGGATAATCGA 2206

```

QY 701 GlyLeuSerArgThr 705
Db 2207 GCCTTGTCAGGACA 2221

RESULT 2
HUMBSA
LOCUS
DEFINITION Human putative transmembrane protein precursor (B5) mRNA, complete cds.
ACCESSION L38961
VERSION L38961.1 GI:624703
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2472)
AUTHORS Lissy,N.A., Bellacosa,A., Sonoda,G., Miller,P.D., Jhanwar,S.C. and Testa,J.R.
TITLE Isolation, characterization, and mapping to human chromosome 11q24-25 of a cDNA encoding a highly conserved putative transmembrane protein, TMC
JOURNAL Biochim. Biophys. Acta 1306 (2-3), 137-141 (1996)
MEDLINE 96221283
PUBMED 8634329
FEATURES
source
    1..2472
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
gene
    1..2472
    /gene="B5"
5'UTR
    1..106
    /gene="B5"
CDS
    107..2224
    /gene="B5"
    /note="putative transmembrane protein precursor"
    /codon_start=1
    /protein_id="AAB05994.1"
    /db_xref="GI:624704"
    translation="MTKEFGFLRLSYEKQDTLLKLLLSMAAVLSFSTRFLPAVLRESV
    LREDPVENVETPEFLAEFGYKHFHWDFDDRAWYPLGRILIGTIVGLMITSAALVHV
    LHPHITIDIRNVCVFLAPLESFSTISVITLLKELKDAGLLAAAMTAVVFGIISR
    SVAGYNDEGAIFCMULTTYIMMKAIVGTGSCICAAKCALAFYVWSSWGGVFFLINL
    IPLVLMITGRSHRIYVAYCTVCLGTLISLRQISFVGFQFVLSSEHMGVFGFL
    CQIHAFVDYLRSKLPQDFEVLRSVTLGVLLTVGALLMTLGTISWTRGFYSLL
    DPSYAKNNIPIIASVSRHQPTWSSVYFDQLLVFMPVGLVYCFENLSDARIFIMY
    GVTSMYSAMVRLMLVLAIPVMSILGIVSOVLSTPMKNLDISRPDKSKKQOOSTY
    PIKIEVASGMILVMAFFLIYTFHSTWVTSEAYSSPSIVLSARGGDSGLIIFDDPERA
    YXMRHNTPEDAKVMKWDYGYQITAMARTILYDNTNWNTHISRVGGMASSTPEKA
    YEIMRELDVSVLIFGLIGYSDDLNKFLMWVRIGGSDTGKHKEKNDYIYTFEYF
    RVDREGSPVLLNCLMYKCYFRFGQVYTEAKRPFPFDRVNAEIGNKDFELDVLEEGY
    TTEHLVRIYKVKDLNRLSRT"
sig_peptide
    107..196
    /gene="B5"
mat_peptide
    197..2221
    /gene="B5"
misc_feature
    452..459
    /product="putative transmembrane protein"
    /gene="B5"
misc_feature
    533..580
    /note="encodes putative transmembrane domain"
    /gene="B5"
misc_feature
    620..667
    /note="encodes putative transmembrane domain"
    /gene="B5"
misc_feature
    746..793
    /note="encodes putative transmembrane domain"
    /gene="B5"
misc_feature
    815..862
    /note="encodes putative transmembrane domain"
    /gene="B5"

```

```

/notes="encodes putative transmembrane domain"
1013..1060
/gene="B5"
/notes="encodes putative transmembrane domain"
1199..1246
/gene="B5"
/notes="encodes putative transmembrane domain"
1310..1357
/gene="B5"
/notes="encodes putative transmembrane domain"
1475..1522
/gene="B5"
/notes="encodes putative transmembrane domain"
2223..2472
/gene="B5"
2455
/gene="B5"
polyA_site
ORIGIN
Alignment Scores:
Pred. No.: 2,566-311 Length: 2472
Score: 3697.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-10-028-384-12 (1-705) x HUMBSA (1-2472)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 107 ATGACTAAGTTTGGATTTTTCGATTTGCTCTATGAGAAGCAGGACACACTTTTGAAGCTT 166
QY 21 LeuLeuLeuSerMetAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGTCATGGCTGCTGATTATCTCTCCACTCGTCTGTTCTGTCCTGAG 226
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTTCAAGAGTTTATCCATGATGTTGATCCGTACTTTTAAATATCGGACTACCAAGTTCTG 286
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaThrTyrProLeu 80
Db 287 GCTGAGAGGGGTTTTATAAATCCATACTGTTTATGACCGAGCCTGGTACCCCTTG 346
QY 81 GlyArgIleLeuGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGACCAATCATGAGGAGAACAAATTTACCCAGGTTTAATGATCACTCTCTGCAATCTAC 406
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTTCACATCACCATCGCATTCGGAAATGCTGTTGTTCTTGCC 466
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 526
QY 141 AlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
Db 527 GCAGGGCTGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 586
QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGTGGCTGGCTCTCTATGATAATGAAGGATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 646
QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGAATCAAGGAGCAAGATGTTTCCATCTGTTGGGACGCTAAGTGTGCCCTT 706
QY 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuLeuLeuLeuPro 220
Db 707 GCTTATTCTTACATGCTCTGCTCATGGGAGGTTATGTGTCTCTGATCACTTAATTTCTCT 766

```

QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 DB 767 CTCACGTCCTCGTGTGATGCTCACAGCGCGTTCTCTCCACCGATCTATGTGGCTAC 826
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 DB 827 TGTACTGTTTACTGCGCTGGGTACTATCTCTTTCAGCAGATCTCTCTTGTGGGTTTCAG 886
 QY 261 ProValLeuSerSerGlnHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 DB 887 CCGTCTCTTCATCAGACACATGGCAGGGTTTGGGCTTTGGTCTCTCCAGATCCAT 946
 QY 281 AlaPheValAspTyrIleuArgSerIysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 DB 947 GCCTTTGTGGATTACTCGCGCAGCAAGTTGAATCCACACAATTTGAAGTCTTTTCCGG 1006
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuMetLeuThr 320
 DB 1007 AGCGTCATCTCTGTGTAGGCTTTGTCTCTCTCACCGTGGAGCTCTCTCATGTGACA 1066
 QY 321 GlyIysIleSerProTyrPheThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaIys 340
 DB 1067 GGAAATAATATCTCCCTGGACGGCGGCTTCTACTCAGTCTGGATCCCTCTTATGCTAAG 1126
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 DB 1127 AACAAATCCCATCTGCTGTGTCTGAGCATGAGCCACAACTGGTCTCTCATAC 1186
 QY 361 TyrPheAspLeuGlnIleuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 DB 1187 TATTTTACCTGAGCTCTCTGCTCTCATGTTTCCAGTTGGGCTCTATATCTGTTAGC 1246
 QY 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1247 AACCTGTCTGATGCCGGATTTTATCATCATGATGTTGTGACCGATGTTACTTTCA 1306
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1307 GCTGTAATGGTGGCTTAATGCTAGTGTGGACCTGTTATGAGCATCTCTCTGGCAT 1366
 QY 421 GlyValSerGlnValLeuSerThrTyrMetIysAsnLeuAspIleSerArgProAspIys 440
 DB 1367 GGAGTCTCCAGGTCGTCTCCATCATCATGAGATCTGGACATTAAGTCGCCAGACAAG 1426
 QY 441 LysSerIysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 DB 1427 AAGAGCAAGCAAGCAAGCAAGTCCACCTACCTATTAAGATTGAAGTGGCAGTGGGATG 1486
 QY 461 IleLeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 DB 1487 ATACTGTCTGATGCTTTCTTCTCATCATCATGCTTCACTCAACCTGGGTGACCAAT 1546
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 DB 1547 GAGGCCCTACTCTCTCCGTCATGTACTATCTGCCCGGTGGGGATGGCAGTAGGATC 1606
 QY 501 IlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 DB 1607 ATATTCATGACTCCGAGAAGCATATATTGGCTTCGTTCATTAATATCTCCAGAGGATCG 1666
 QY 521 LysValMetSerTrpThrPheTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 DB 1667 AAGTCTATGCTCGGTGGGATTAAGCTATCAGATTACAGCTATGGCAACCGCAACAT 1726
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1727 TTAGTGACAATAACATGGAATAATATCCATATTTTCGAGTAGGCGAGGCAATGGCG 1786
 QY 561 SerThrGluGluIysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1787 TCCACAGAGAAAGGCTATGAGATCATGAGGAGCTCGATGCTGAGCTATGTGTGCTG 1846
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600

DB 1847 ATTTTGGAGGCTCCTCTGATGATATCAACAGTTCTTCTGGATGGTC 1906
 QY 601 ArgIleGlyGlySerThrAspThrGlyHisIleLysGluAsnAspTyrTyrThrPro 620
 DB 1907 CGGATGGAGGAGCAGATACAGGCAACATATCAAGGAGATGACTATATATCTCA 1966
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 DB 1967 ACTGGGAGTTCCGTTGTGGACCGTGAAGGTTCTCAGTGTCTCAACTGCTCATGTAC 2026
 QY 641 LysMetCysTyrTyrArgPheGlyValTyrThrGluAlaLysArgProGlyPhe 660
 DB 2027 AAGATGTTTACTATCGCTTTGGACAGTTTACAGAGCCAGGCTCTCCAGGCTTT 2086
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
 DB 2087 GACCGTGTCCGAAATGCTGAGATTGGGAATAAGACTTTGAGCTTGTCTCGGAGAA 2146
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 DB 2147 GCTATATACAGAACATTTGCTGTCAGGATATACAGGTAAAGGACCTGGATATATCGA 2206
 QY 701 GlyLeuSerArgThr 705
 DB 2207 GGCTTGTCAAGGACA 2221
 RESULT 3
 BT007100 2118 bp mRNA linear PRI 13-MAY-2003
 LOCUS Homo sapiens integral membrane protein 1 mRNA, complete cds.
 DEFINITION
 ACCESSION BT007100
 VERSION BT007100.1 GI:30583038
 KEYWORDS FLI CDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2118)
 Kalline, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S.,
 Koundinya, M., Raphael, J., Moreira, D., Kelley, T., Labaer, J., Lin, Y.,
 Phelan, M. and Farmer, A.
 Cloning of human full-length CDSs in BD Creator(TM) System Donor
 vector
 Unpublished
 2 (bases 1 to 2118)
 Kalline, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S.,
 Koundinya, M., Raphael, J., Moreira, D., Kelley, T., Labaer, J., Lin, Y.,
 Phelan, M. and Farmer, A.
 Direct Submission
 Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
 Circle, Palo Alto, CA 94303, USA
 This CDS clone is a part of a collection of human full length
 expression clones generated by BD Biosciences Clontech and the
 Harvard Institute of Proteomics. Each CDS has been cloned in two
 forms: with and without stop-codon (to allow fusion with C-terminal
 tag). The CDS has been directionally cloned using BD in-Fusion(TM)
 cloning system between the SalI and HindIII sites of the pDNR-DUAL
 vector. Additional sequences in the clone: 'ACC' after SalI site
 and before 'ATG' to provide Kozak consensus sequence; 'GG' after
 last codon and before HindIII site to maintain reading frame.
 Clone distribution: http://bioinfo.clontech.com/orfclones.
 location/Qualifiers
 1. 2118
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GH00707X1.0"
 /clone.lib="BD Creator(TM) CDS Library derived from MGC
 collection"
 /lab_host="DH5alpha T1 resistant"
 /note="vector: pDNR-Dual"
 1. 2118
 CDS


```

/codon_start=1
/product="integral membrane protein 1"
/protein_id="AAP35764.1"
/db_xref="GI:30583039"
/translation="MTKEGFLRLSEVEKQDTLLKLLLSMAAVLSFSTRFLPAVRFSEV
IHEPDVENVETTFLEAEFGYKPHNFDDRWLPLGRIIGGTYIPLGMITSAIYHV
LHPHIIIDIRNVCVFLAPLFFSSITIVYHLTKELDAGALLAAIAVAPGYISR
SVAGYDNEGAIFCMULTYIMWKTKGSIQWAAKCALAYFYVMSWGWYVFLINL
IPLHVLMLTRFSHR IYAACTVYCLGILSMQISFVGFQVPLSSEHMAAFVFLG
COIHFVDYLRKLNPOQFVLFPSVSLVFLVLLTVGALLMLTGKISPMWGRFSYL
DPSYAKNNIPPIAASVBHOPTWSSYVFDLQVFMPPVGLVYCFNLSDPARPIIMY
GVTSMYSAVMRLMLVLPVWCILSGIGYVSVLSTYMKNLDIRDCKSKQKQDSTY
PIKNEVASMTLWAFELITYTHSTWITSEAYSSPISVLSARGGDSRIIPDFREA
YYLNRHNPEDAKKSWMDYQITAVNATILVDNNTWNTHSIRVGQAMASTEKA
YEINRELDVSLVLFGLTGYSDDLNKLFWIRGGSTDTGKHENDYYTPTEF
RVDREGSPVLLNCLMXYRFGQVYTEAKRPPGDFRVERNAIGNKDFELDLVEAY
TTEHLVRIYKVKDLNRGLSRT"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 6 2e-308 Length: 2118
Score: 3657.00 Matches: 698
Percent Similarity: 99.15% Conservative: 1
Best Local Similarity: 99.01% Mismatches: 6
Query Match: 98.92% Indels: 0
DB: 9 Gaps: 0

```

US-10-028-384-12 (1-705) x BT007100 (1-2118)

```

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
DB 1 ATGACTAAGTTTGGATTTCGCGATTGCTCTATGAGAAGCAGACACACTTTTGAAGCTT 60

QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 61 CTGATTCGTCATGGCTGCTGATTATCTCTCCACCTGCTGTTGCTGCTGCTGCTGCTG 120

QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
DB 121 TTGTAAAGTGTATTCATGAGTTTGTATCCGCTACTTTAATATCGGACTACCGAGTTCC 180

QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTyrProLeu 80
DB 181 GCTGAGAGAGGGGTTTATAAATCCATACTGTTGATGACCGAGCCTGGTACCCCTTTG 240

QY 81 GlyArgIleileGlyGlyThrIleTyrProGlyLeuMetIleAlaValProGlyTyrIleSerArg 100
DB 241 GGACGAATCATTCGAGGACAAATTTACCAAGGTTTAAATGATCACCTCTCTGCAATCTAC 300

QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
DB 301 CATGTACTCCATTTTTCACATCATCACATCGACATTCGGAATGTCGTGTGCTGCTGCTG 360

QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
DB 361 CCTCTCTTCTCTCTCTCCACCATTCGTCAGTCAGTACCATCTTACCAAGAGCTCAAGGAT 420

QY 141 AlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
DB 421 GCAGGGGCTGGGCTTCTGCTGCTGCATGATTGCTGTAGTTCTCTGGATATATCTCCCGA 480

QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
DB 481 TGTGTGGCTGGCTCTCTATGATATGAAGGGATTGCGATCTTTTGTGATGCTACTACCTTAC 540

QY 181 TyrMetTrrPileLysAlaValLysThrGlySerIleCysTrrPAlaAlaLysCysAlaLeu 200
DB 541 TACATGTGGATCAAGGACGATAAGACTGGTTCATCTGTTGGGAGCTAAGTGGCCCTT 600

QY 201 AlaTyrPheTyrMetValSerSerTrrPglyGlyTyrValPheLeuLeuAsnLeuPro 220
DB 601 GCTTATTCTACATGGCTCTCGTCATGGGAGGTTATGTGTTCTCTGATCAACTTAATCTCT 660

```

```

QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
DB 661 CTCACGTCCTCGTGTGATGTACAGGCCCTTTCTCTCACCGGATCTATGTGCGCTAC 720

QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
DB 721 TGTACTGTTTACTGCTGGGCACTATACCTTTCTATGCAGATCTCTTTGTGGTTTCCAG 780

QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
DB 781 CTGTGCTTTTCATCAGAGCACATGCGACCTTTTGGGCTCTTTGGTCTCTGCGCAGATCCAT 840

QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
DB 841 GCCTTTTGGAATTACTCGCGCAGCAGTGGATCCACACAACTTTGAAGTCTTTTCCGG 900

QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
DB 901 AGCGTCATCTCTCGTGTAGGCTTTGCTCTCTCACCGTGGGAGCTCTCTCATGCTGACA 960

QY 321 GlyLysIleSerProTrrPThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
DB 961 GGAAAAATATCTCCCTGACGGGGGCTTTCTACTCGCTGCTGGATCCCTCTTATGCTAAG 1020

QY 341 AsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrrPserSerTyr 360
DB 1021 AACAAACATCCCATCATTTGCTGTCTGAGCATCAGCCACAACTGCTGCTCTCATAC 1080

QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
DB 1081 TATTTTGACCTGCGAGCTCTCTGCTTCAAGTTTCCAGTTTCCAGTTGCGCTCTATTA 1140

QY 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
DB 1141 AACCTGTCTGATGCCGGATTTTATCATCATGATATGCTGTGACGACCATGATCTTTTCA 1200

QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
DB 1201 GCTGTAAATGGTGGCTGTAAATGCTAGTGTGGCACCTGTTATGTGCATTTCTCTCTGG 1260

QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
DB 1261 GGAGTCTCCAGGTGCTGCCACATACATGAGAAATCTGGACATAGTCTGCCAGACAAG 1320

QY 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
DB 1321 AAGAGCAAGAACAGCAACAGGATTCACCTACCTATTAAAGAAATGAAGTGGCAAGTGG 1380

QY 461 IleLeuValMetAlaPhePheLeuLeuThrTyrPheHisSerThrTrrPValThrSer 480
DB 1381 ATACTGGTCAATGGCTTTCTTTCTCATCACCTTACACCTTTTATTCAACCTGGGTGAC 1440

QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
DB 1441 GAGGCTTACTCTCTCGGTCATGTAATGTAATCTATCTGCCGCTGGTGGGATGTCAGTAG 1500

QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrrPleuArgHisAsnThrProGluAspAla 520
DB 1501 ATATTGATGACTTCCGAGAGCATATTATTGGCTTCTGCTCATATACTCCAGAGGATGG 1560

QY 521 LysValMetSerTrrPrrPAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
DB 1561 AAGGTCATGCTCTGGTGGGATTAATGGCTATCAGATTTACAGCTATGGCAACCCGACAA 1620

QY 541 LeuValAspAsnAsnThrTrrPAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
DB 1621 TTAGTGGACAAATACACATAGGAATAATACCAATATTCTTCTGAGTAGGAGGCAATGG 1680

QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTrrValLeuVal 580
DB 1681 TCCACAGAGGAAAAGCCATATGAGATCATGAGGAGGCTCGATGTCAGCTATGTGCTGGTC 1740

QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600

```

```

1741 ATTTTGGAGCCCTCAGTGGGTATCTCTGATGATATCAACAGTTCTTTGGATGTC 1800
QY 601 ArgIleGlySerThrAspThrGlyHisIleGlyLeuAspTyrThrPro 620
DB 1801 CGGATTGGAGGACACAGATACAGGCAACATATCAAGGAGATGATATATATCTCA 1860
QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
DB 1861 ACTGGGAGATTCCTGTGGACCGGTGAAGTTCTCCAGTGTCTCAACTGCCTCAATGAC 1920
QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
DB 1921 AAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAGCAAGCCCTCCAGGCTTT 1980
QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
DB 1981 GACGTGTCCGAATGCTGAGATTGGATAAGACTTTGAGCTTGATGCTCTGGAGAA 2040
QY 681 GlyTyrThrThrGluHisTyrPheValArgIleTyrLysValLysAspLeuAspAsnArg 700
DB 2041 GCATATACACAGAACATTCGCTGTCAGCATATACAGGTAAGAGCACTCGATATCGA 2100
QY 701 GlyLeuSerArgThr 705
DB 2101 GGCTTGTCAAGGACA 2115

RESULT 4
BT008132
LOCUS
DEFINITION Synthetic construct Homo sapiens integral membrane protein 1 mRNA,
partial cds.
ACCESSION BT008132
VERSION BT008132.1
KEYWORDS GLI:30585102
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 2118)
AUTHORS Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator (TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2118)
AUTHORS Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion (TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
source
1..2118
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone_lib="BD Creator (TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/notes="Vector: pDNR-Dual"
1..>2118
CDS

```

```

/notes="Mutations: 2117:Stop->Leu"
/codon_start=1
/transl_table=11
/product="Homo sapiens integral membrane protein 1"
/protein_id="AAP36824.1"
/db_xref="GI:30585103"
/translation="MTKPGFLRLSYEKQDTLLKLLLSMAAVLSFSTRLFAVLPFESV
LHEPDYNYTTRFLAREGFKPHNFDDRAWYPLGRIIGGTIYPLGMITSAIYHV
LHPHTIDIDNVCEFLAPLSPSTTIYVHLKELKADAGALLAAVAVPGYLSR
SVAGSYDEGAIACMLITYNMKIKAVTSGICWAACALAYFWVSSKGGYVFLNL
LPLHVLMDLFRSHRIYVAYCTVYCIIGTILSMQISFVGFQVLSERMAAFVGL
COIHAFVLRSLNPOQFVLRFSVISLVGFVLLTVGALLMTGKISPMWTFGRFLL
DPSAKNNPIIASVSEHQPTTWSSYFDLQLLVFMPVGLYFCFNLSDARFIYLL
GVTMYFGAVMRLMLVLPWCILSGIVSOVLSTYMKLDIRDPDKSKKKQDSTY
PKNEVASGMILVNAFLITYVHTSTWVTSYVSSPSIVLSARGGSGSRIIFDPPEA
YVLRHNTPEDAKVMWDYQITAMANTILVNNNTIIVNNTIISVAGQMASTBKA
YEIRLEDSYLVLTFGLTGYSSDDINKFLMWVRLGGSTDTGCKHKENDYVTPTEF
RVDRGSPVLLNCLMYKCYRFGQVYTEAKRPPGPDVRVNAISGNKDFELDLVLEAY
TTEHVLVRIYKVDLNRGUSRII"
Alignment Scores:
Pred. No.: 6,2e-308 Length: 2118
Score: 3657.00 Matches: 698
Percent Similarity: 99.15% Conservative: 1
Best Local Similarity: 99.01% Mismatches: 6
Query Match: 98.92% Indels: 0
DB: 12 Gaps: 0
US-10-028-384-12 (1-705) x BT008132 (1-2118)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
DB 1 ATGACTAGATTGGATTTTTCGATTGTCCTATGAGAGGACGACACACTTTTGAAGCTT 60
QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 61 CTCATTCTGCAATGGCTGCTGATTATCTCTCCACTGCTGCTGTTTGTCTGCTGAGA 120
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
DB 121 TTGGAAGTGTATCCATGAGTTGATCCGACTTTAATATCGGACTACACAGGTTCTG 180
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaIleTyrProLeu 80
DB 181 GCTGAGGAGGGGTTTATAAATTCATAACTGGTTTGTATGACCGAGCCTGCTACCTTG 240
QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 241 GGACGAATCATGGAGGACAAATTATCCAGGTTTATGATCATCTCTGCTGCAATCTAC 300
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
DB 301 CATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGCTGCTGTTGTTCTGGCC 360
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
DB 361 CTCCTCTCTCTCTCTCCCTCACCACCATCGTCACGTAACCACTTACCAGAGCTCAAGAT 420
QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
DB 421 GCAGGGGCTGGGCTCTTTGCTGCTGCCATGATTCCTAGTTCCTGATATATCTCCCGA 480
QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
DB 481 TCTGTGGCTGGCTCCTATGATAATGAGGAGATTCCTCTTTGATGCTACTACCTAC 540
QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
DB 541 TACATGTGGATCAAGGACGTAAGACTGGTTCCTATCTGTTGGGACGAGTAAGTGTGCCCT 600
QY 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuLeuLeuLeuPro 220

```

Db 601 GCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTGTCTCTGATCAACTTAATTCCT 660
 Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 Db 661 CTCACAGTCTCTGCTGCTGATGCTCACAGGCGTTTCTCTCACGGATCTATGTGGGCTAC 720
 Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db 721 TGTACTGTATTACTGCTGGGCACTATATCTTTCTATGCAGATCTCTCTTGTGGGTTTCCAG 780
 Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 Db 781 CTTGTCCTTTTATCAGACCATATGGCAGCCTTTGGGCTCTTGGTCTCTGCCAGATCCAT 840
 Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db 841 GCCTTTGTGGATTACTCTGGCAGCAAGTTGAATCCACAACTTGAAGTCTCTTTTCCGG 900
 Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 Db 901 AGCGTCATCTCTGTGTAGGTTTGTCTCTCTCACCGTGGAGCTCTCTCTCATGCTGACA 960
 Qy 321 GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 Db 961 GCAAAATATCTCCCTGGACGGGCGTTTCTACTCGCTGCTGGATCCCTCTTATGCTAAG 1020
 Qy 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrSerSerTyr 360
 Db 1021 AACACATCCCATCATTTCTCTGTGTGAGGATCAGCCCAACCTGGTCTCTCATAC 1080
 Qy 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 Db 1081 TATTTTGCACCTGCT 1140
 Qy 381 AsnLeuSerAspAlaArgPheIleLeuMetTyrGlyValThrSerMetTyrPheSer 400
 Db 1141 AACCTGTCTGATGCCGGAATTTTATCATCATGATGGTGACCAAGCATGTACTTTCA 1200
 Qy 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 Db 1201 GCTGTAAATGGTGGCTTAATGCTAGTGTGGACCTGTATTGTGCTATCTCTCTGGCAT 1260
 Qy 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 Db 1261 GGAGTCTCCAGTCTCTCTCACATCATATGAAGATCTGGACATTAAGTCTCTCACAGCAAG 1320
 Qy 441 LysSerLysLysGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 Db 1321 AAGAGCAAGACAAACAGGATTCACCTACCTTAAAGATGAAGTGGCAAGTGGGATG 1380
 Qy 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTyrValThrSer 480
 Db 1381 ATACTGGTCATGGCTTTCTCTCTCATCATCATCATCATCATCATCATCATCATCATCAT 1440
 Qy 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 Db 1441 GAGGCGTACTCTCTCCGTCCTATCTATCTGCTCCGCTGGTGGGATGGCAGTAGGATC 1500
 Qy 501 IlePheAspPheArgGluAlaTyrTyrTyrProLeuArgHisAsnThrProGluAspAla 520
 Db 1501 ATATTGTAGTATCCGAGAGCATATTATTGGTCTGTCTATATATCTCAGAGGATGGC 1560
 Qy 521 LysValMetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 Db 1561 AAGGTCATGCTCTGGTGGATTTATGGCTATCAGATTAAGCTATGGCAAAACCGAACAT 1620
 Qy 541 LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1621 TTATGTGGCAATPACACATGGATGGAATATATCCCATATTTCTCAGTAGGCGAGCATGGC 1680
 Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1681 TCCACAGAGCAAAAGCCCTATCAGATCATGAGGAGCTCGATGTGCTATGTCTGGTCT 1740

Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTyrMetVal 600
 Db 1741 ATTTTGGAGGCTCTACTGGTATCTCTCTGATGATATCAACAAGTTTCTTTGGATGGTC 1800
 Qy 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1801 CGGATTGGAGGGAGCAGATACAGCAACATATCAAGGAGATGACTATTATTAATCTCA 1860
 Qy 621 ThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1861 ACTGGGAGTTCCGTGTGGACCGTGAAGGTTTCTCCAGTCTGCTCAACTGCCTCATGTAC 1920
 Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 1921 AAGATGTGTACTATCTCTTGGACAGGTTTACACAGAGCCAGGCTCTCCAGGCTTT 1980
 Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 1981 GACCGTGTCCGAAATGCTGAGATTGGGATATAAGACTTTTGAGCTTCATGCTCTGGAGAA 2040
 Qy 681 GlyTyrThrThrGluHisTyrProLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2041 GCATATACACAGAACATTTGGCTGGTTCAGGATATACAGTAAGGACTTGATATACTGA 2100
 Qy 701 GlyLeuSerArgThr 705
 Db 2101 GGCTTGTCAAGGACA 2115
 RESULT 5
 BC048348
 LOCUS
 DEFINITION Homo sapiens integral membrane protein 1, mRNA (cdna clone IMAGE:5561615), partial cds.
 ACCESSION BC048348
 VERSION BC048348.1
 KEYWORDS GI:29477243
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2450)
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Young, A.C., Shevchenko, Y., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Burtchfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Scheraga, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 2450)
 Strausberg, R.
 Direct Submission
 Submitted (07-MAR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk

REMARK
COMMENT

Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nri.nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dierich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Mastelco, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantropop, S., Thomas, P.J., Touchman, J.W.,
Tsurgueon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 93 Row: P Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 22749414.

FEATURES

Location/Qualifiers
1..2450
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5561615"
/tissue_type="Duodenum, adenocarcinoma"
/clone_lib="NIH MGC_88"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..2450
/gene="ITM1"
/note="synonyms: TMC, MGC9042"
/db_xref="LocusID:3703"
/db_xref="MIM:601134"
1..2163
/gene="ITM1"
/codon_start=1
/product="ITM1 protein"
/protein_id="AAH48348.1"
/db_xref="GI:29477244"
/db_xref="LocusID:3703"
/db_xref="MIM:601134"

gene

CDS

ORIGIN

Alignment Scores:
Pred. No.: 7.59e-308 Length: 2450
Score: 3657.00 Matches: 698
Percent Similarity: 99.15% Conservative: 1
Best Local Similarity: 99.01% Mismatches: 6
Query Match: 98.92% Indels: 0
DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x BC048348 (1-2450)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeu 20
|||||

Db 46 ATGACTAAAGTTGGATTTTTCGGATTGTCTATCAGAAAGCAGGACACACTTTTGAAGCTT 105
Qy 21 LeuLleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
|||||
Db 106 CTCATCTCTGTAATGGCTGCTGTATTATTCCTTCTCCACTCTGCTGTTGCTGCTCAGAGA 165
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
166 TTTGAAAGTGTTATCCATGAGTTTGGATCCGTTATTTAATTATCGGACTTACCAGGTTCCTG 225
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
226 GCTGAGGAGGGGTTTATAAATTCATAACTGGTTTGAAGCAGGCTGTGTACCTTTG 285
Qy 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
286 GGACGAAATCATTTGGAGGAACTATTACCAGGTTTATGATCACCCTCTGCTGCAATCTAC 345
Db 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
346 CATGTACTCCATTTTTCACATCCATCGACATTCGGAATGTCTGTGTCTCTGTCGCC 405
Qy 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
406 CCTCTCTTCT 465
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
466 GCAGGGGCTGGGCTCTTCTGCTGCTCATGATTCGTGTAGTTCTCTGGATATATCTCCCGA 525
Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
526 TCTGTGCTGCTGCT 585
Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
586 TACATGTGGATCAAGGACGTAAGAGCTGGTTCATCTCTGTGGGACGTAAGTGTGCCCTT 645
Qy 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuLysLeuLysPro 220
646 GCTTATTCTTACATGCT 705
Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
706 CTCACAGTCTCTGCTGCTGATGCTCAGAGGCGCTTCTCTCTCACGGATCTATGTGGCTAC 765
Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
766 TGTACTGTTTACGCTCTGGGCACTATATCTCTATGTCAGATCTCTCTTGTGGGTTCACAG 825
Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
826 CCTGTCTCTTTCATCAGAGCAGATGGGAGCTTTGGGCTCTTTGGTCTCTGCGAGATCCAT 885
Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArg 300
886 GCCTTTCTGGATTACCTGCGCAGCAAGTTGAATCCACAATTTTGAAGTCTTTTTCGCG 945
Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyValAlaLeuLeuMetLeuThr 320
946 AGCGTCACTCTCTCTGTTAGGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 1005
Qy 321 GlyLysIleSerProTyrThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
1006 GGAAATAATCTCCCTGGAGCGGGCGTTTCTACTCGTGTGATGATCCCTCTTATGTGAAG 1065
Qy 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
1066 AACAAACATCCCATCATTTGCTTCTGTGCTGAGCATCAGCCCAACACCTGGTCTCTATAC 1125
Qy 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
1126 TATTTTGCCTGAGCTCTCTCGTCTTCTATGTTTCCAGTTTGGCTCTATTACTGCTTTAGC 1185
Db

QY	361	AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer	400
DB	1186	AACTGTCGTGATGCCCGGATTTTATCATCATGTATGGTGCACGAGATGACTTTTCA	1245
QY	401	AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle	420
DB	1246	GCTGTATAGTGGCGCTAAATGCTAGTGTGGCACCTGTTATGTGCAATCTCTCTGGCATT	1305
QY	421	GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys	440
DB	1306	GGAGTCTCCCAAGTGCTGTCCACATACATGAAGAATCTGGACATAAAGTCGTCACAGACAAG	1365
QY	441	LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet	460
DB	1366	AAGAGCAAGACCAACAGGATTCACCTACCCCTATTAAAGATCAAGTGGCAAGTGGGATG	1425
QY	461	IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer	480
DB	1426	ATACGGTTCATGGGCTTTCTTCTTCATCCCTACCTTCATCCCTTCATTCAACTGGGTGACCAAGT	1485
QY	481	GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle	500
DB	1486	GAGGCGCTACTCTCTCCGTCCTATTGCTACTATCTGCCGCTGGTGGGATGGCAGTAGGATC	1545
QY	501	IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla	520
DB	1546	ATATTGTGATGCTCCGAGGAACATATTATGGCTTCGTCATATPACTCCAGAGGATGCG	1605
QY	521	LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle	540
DB	1606	AAGGTCATGTCCTCGTGGGATTTATGGCTATCAGATTACAGCTATGGCAACCGAACCAATT	1665
QY	541	LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla	560
DB	1666	TTATGTGACATACACATGGAAATATACCATAATTTCTCGAGTAGGCGAGGCATGGCG	1725
QY	561	SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal	580
DB	1726	TCCACAGAGGAAAAGCGCTATGAGATCATGAGGAGCGTCGATGTCAGGTATGTGCTGGTC	1785
QY	581	IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal	600
DB	1786	ATTTTGGAGCGCTCACTGGGTATCTCTGTGATGATATCAACAAGTTTCTTTGGATGGTC	1845
QY	601	ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro	620
DB	1846	CGGATTGGAGGAGCAGATACAGCAACATATCAAGGAGAATGACTATATATACTCCA	1905
QY	621	ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr	640
DB	1906	ACTGGGGAGTTCCGTTGTGGACCGTGAAGGTTCTCCAGTCTGCTCAACTGCCTCATGTAC	1965
QY	641	LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe	660
DB	1966	AAGATGTGTTACTATCCGTTTGACAGGTTTACACAGAAGCCAAAGCGTCTCTCCAGGCTTT	2025
QY	661	AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu	680
DB	2026	GACCGTGCOCGAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGTCTCGTGAGAA	2085
QY	681	GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspArg	700
DB	2086	GCATATACACAGAACATTTGGCTGTCAGGATATACAGGTAAAGGACTGGTAATCGA	2145
QY	701	GlyLeuSerArgThr	705
DB	2146	GGCTGTGCAAGGACA	2160

RESULT	LOCUS	DEFINITION
6	CQ834714	

linear PAT 29-JUL-2004

ACCESSION	CQ834714
VERSION	CQ834714.1
KEYWORDS	GI:50834251
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1
AUTHORS	Matsuda,A. and Yoneta,S.
TITLE	T cell activating gene
JOURNAL	Patent; WO 2004058805-A 585 15-JUL-2004;
FEATURES	Asahi Kasei Pharma Corporation (JP) Location/Qualifiers 1..2456 /organism="Homo sapiens" T cell activating gene /mol_type="unassigned DNA" /db_xref="taxon:9606" 93..2210 /note="unnamed protein product" /codon_start=1 /protein_id="CAH05572.1" /db_xref="GI:50834252" /translation="MTKFGFLSLSEYKODTLKKLLLSMAVLFSFTLRFAVLFRRFSSV IHEPDPENYVTRTFRLAEAGGFVKFNHEDDRAWPLGRIIGLGTIPGLMITSAAIVHY LHPHIIDIRNCVCLALBLRSPFITLYTHLTRELKDAGALLAAMIAKVGPYSIR SVAGSDNEGIALFCMLLIYHWIKAVXIGTSIQWISQICILSMOISFVGQPVLISEHHAARFVGLNL IPLVHLNMGTFSHRIIVAYCTVCIGTILSMQISFVGQPVLISEHHAARFVGLNL COHFAPVDITLCNQPEQLFVRYSILVSGEVLITVGAALLMTGKLSPWTGRFPIYL DPSYAKNIIDIIASVEHQPTWSSYFDQLLFVMFPVGLHYCFNLSDARFIYML DTGSYFSAVMRLMLAFVNMCIISLGIVGCVLSITYMNLDISRPOKSKKQQDSTY PIKEVASGMILWMAPFLITYFHSTWTSYAYSPPSIVLSARGDGSRIFDDFREAA YYLMRLNPDAKMWSMDYQGQITAMARNILVNNTNNTHISRVGVQAMASTEBKA YEIMRELDVSYLVIIEGGTLTGSDINKFLMWVRIGGSTDTGKHILENDVYVTPGFR RVDREGSPVLLNCLMKVCYIRFGQGYTEAKPGPFDRVNAIEGNKDFELDLVELEY TTETHWLVIYKYVDLNRLGSLRT"
source	
CDS	
ORIGIN	
Alignment Scores:	
Pred. No.:	7,61e-308
Score:	3657.00
Percent Similarity:	99.15%
Best Local Similarity:	99.01%
Query Match:	98.92%
DB:	6
	Gaps: 0
	Indels: 0
	Mismatches: 6
	Conservative: 1
	Matches: 698
	Length: 2458

```

US-10-028-384-12 (1-705) x CQ834714 (1-2458)

QY      1 MetThrLysPheGlyPheLeuA-gLeuSerTyrgLUySGInAspThrLeuLeuLysLeu 20
Db      93 ATGACTAAGTTTGGATTTTGGCATTTGCTCATGAGACGACGACACACTTTTGAAGCTT 152
QY      21 LeulleLeuSerMetAlaIaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db      153 CTCATTCCTGTCAAATGGCTGCTGATTATCCTTCTCCATCGTCTGTTGCTGCTCGAGA 212
QY      41 PheGluSerValIleHisGluPheAspProTyrgPheAsnTyrgThrThrArgPheLeu 60
Db      213 TTTGAAAGCTGTTATCCATGAGTTTGATCCGACTTTAATTATCGGACATCCAGAGTTCCTG 272
QY      61 AlaGluGluGlyPheTyrgLysPheHisAsnTrpPheAspAspArgAlaTrpTyrgProleu 80
Db      273 GCTGAGGAGGGGTTTTATAAATCCATAACTGGTTGTATGACCGAGCTCGTGACCCCTTG 332
QY      81 GlyArgIleIleGlyGlyThrIleTyrgProGlyLeuMetIleThrSerAlaAlaIleTyrg 100
Db      333 GGACGAATCAITGGAGGAACAATTACCAGGTTTAATGATCACCTCTGCTGCAATCTAC 392
QY      101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db      393 CATGTACTCCATTTTTTCCACATCACCATCGACATTCGGAATGTCTGTGTCTTCTCGGCC 452
QY      121 ProLeuPheSerPheThrSerIleValThrTyrgLeuLeuThrLysGluLeuLysAsp 140

```

453	DB	CCTCTCTTCTCTCTCTTCCACCCACCATCGTCGATNCCACCTTACCAAGAGAGCTCAAGGAT	512
141	QY	AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrrIleSerArg	160
513	DB	CGAGGGCGTGGCGTCTTGTGCTGCGCATGATGCTGTAGTTCTGTGGATATATCTCCCA	572
161	QY	SerValAlaGlySerTyrrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrr	180
573	DB	TCTGTGCTGGCTCCTATGATATATGAAGGATTTGCCATCTTTTGCATGCTACTCACCTAC	632
181	QY	TyrrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu	200
633	DB	TACATGTGGATCAAGGCAGTAAAGACTGGTTCATCTGTGTGGCAGCTAAGTGTGCCCTT	692
201	QY	AlaTyrrPheTyrrMetValSerSerTrpGlyGlyTyrrValPheLeuIleAsnLeuIlePro	220
693	DB	GCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTTCCTGATCACTTAATTCCT	752
221	QY	LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrrValAlaTyrr	240
753	DB	CTCCACGTCCTCGTGTGTGATGCTCAAGACCCGCTTTCTCTCACCGGATCTATGTGGCTAC	812
241	QY	CysThrValTyrrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln	260
813	DB	TGTACTGTTTACTGCTGGGCACCTATCTTCTATGCAGATCTCTTTGTGGGTTTCCAG	872
261	QY	ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis	280
873	DB	CCTGCTCTTTTCATCAGACACATGGCAGCGCTTTGGGGTCTTTGGTCTCTGCCAGATCCAT	932
281	QY	AlaPheValAspTyrrLeuArgSerLysIleuAsnProGlnGlnPheGluValLeuPheArg	300
933	DB	GCCTTTGTGATTAACCTGGCAGCAAGTTGAATCCACAACTTTGAAGTCTCTTTCCCG	992
301	QY	SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr	320
993	DB	AGCGTCATCTCTGTGTAGGCTTTGTCTTCACCGTGGGAGTCTCTCTCATGTGCACA	1052
321	QY	GlyLysIleSerProTrpThrGlyArgPheTyrrSerLeuLeuAspProSerTyrrAlaLys	340
1053	DB	GGAAAAATATCTCCCTGGACGGGGCGTTTCTACTCGTCTGGATCCCTCTTATGCTAAG	1112
341	QY	AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyrr	360
1113	DB	AACAAACATCCCATCATCTGCTGTGTGCTGAGCATCAGCCACAACTGGTCTCTCATAC	1172
361	QY	TyrrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrrTyrrCysPheSer	380
1173	DB	TATTTTGACCTGCAGCTCTCTGCTCTCATGTTCCTTCCAGTTGGCTCTTACTGCTTAGC	1232
381	QY	AsnLeuSerAspAlaArgIlePheIleIleMetTyrrGlyValThrSerMetTyrrPheSer	400
1233	DB	AACCTGTCTGATGCCCGGATTTTATCATCATGTATGTGTGACACAGCATGTACTTTTCA	1292
401	QY	AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle	420
1293	DB	GCTGTAATGTGTGGCTCTAATGCTAGTGTGTGGACCTGTTATGTCACTTCTCTGGCATT	1352
421	QY	GlyValSerGlnValLeuSerThrTyrrMetLysAsnLeuAspIleSerArgProAspLys	440
1353	DB	GGAGCTCCCCAGGTCGTCCACATACATGAAGAATCTGGACATAAGTCGTGCAGACAAG	1412
441	QY	LysSerLysLysGlnGlnAspSerThrTyrrProIleLysIleGluValAlaSerGlyMet	460
1413	DB	AAGAGCAAGAAGCAACAGGATTCACCTACCCTATTAAAGATGAAGTGGCAAGTGGGATG	1472
461	QY	IleLeuValMetAlaPhePheIleuIleThrTyrrPheHisSerThrTrpValThrSer	480
1473	DB	ATATGCTGTCAGGCTTTCTTTCTCATCACTACCTTTCATTCAACTGGGTGGACAGT	1532
481	QY	GluAlaTyrrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle	500
1533	DB	GAGCCCTACTTCTCTCGTCCATTGACTATCTGCCGTGTGGGTGGATGCACTAGATC	1592

501	QY	IllePheAspaspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla	520
502	QY	ATATTTGATGACATTCGAGAACATATTAATTTGGCTTCGTATATATCTCAGAGGATGGC	1652
503	DB		
521	QY	LysValMetSerTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle	540
522	QY	AAAGTCATGTCTCTGGTGGGATTAAGCTATCAGATTACAGCTATGGCAACCGAACAAAT	1712
523	DB		
541	QY	LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla	560
542	QY	TTAGTGGACAATAAACAATGCGAATAATACCATATTTCTCGAGTAGGCGAGGCAATGGCG	1772
543	DB		
561	QY	SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal	580
562	QY	TCCACAGAGGAAAAAGCCATATCAGAGATCATGAGGAGAGCTCGATGTACAGTATGTCTGGTGC	1832
563	DB		
581	QY	IlePheGlyClyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTyrPMetVal	600
582	QY	ATTTTGGAGCGCTCACTGGGTATTCCTCTCATGATATCAACAAGTTTCITTCGTATGGTGC	1892
583	DB		
601	QY	ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro	620
602	QY	CGGATTGGAGGAGACACAGATACAGGCAAAATATCAAGGAGATGATGACTATTATACTCCA	1952
603	DB		
621	QY	ThrClyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr	640
622	QY	ACTGGGGAGTTCCGTGTGACCGTGGAAGGTCTCCAGTGTCTCTCACTGCCTCATGTAC	2012
623	DB		
641	QY	LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe	660
642	QY	AAAGTGTGTTACTATCGCTTTGAGCAGGTATTTACACAGAAGCAAGCGTCTCCAGGCITTT	2072
643	DB		
661	QY	AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu	680
662	QY	GACCGTGTCCGAATGCTCAGATTGGGAATTAAGACITTTGAGCTTGATGTCTCGTAGGAA	2132
663	DB		
681	QY	GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg	700
682	QY	GCATATACACAGAAATTTGGCTGGTCTCAGGATATACAAAGTAAAGGACCTGGGATAATCGA	2192
683	DB		
701	QY	GlyLeuSerArgThr	705
702	QY	GGCTTGTCAAGGACA	2207
703	DB		
RESULT 7			
BC020965	LOCUS		
DEFINITION			
Homo sapiens integral membrane protein 1, mRNA (cdna clone MGC:9042			
IMAGE:3891543), complete cds.			
ACCESSION			
BC020965			
VERSION			
BC020965.2 GI:34190969			
KEYWORDS			
MGC.			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE			
1 (bases 1 to 2516)			
AUTHORS			
Krausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,			
Stauberg,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,			
Altshul,S.F., Zeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,			
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,			
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,			
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,			
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toehyuki,S.,			
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,			
Aranson,R.D., Mullany,S.J., Bosak,S.A., McEwen,P.J.,			
Mokry,K.C., Hale,S.A., Garcia,A.M., Gay,L.J., Hulyk,S.W.,			
Villalon,D.C., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,			
Fahey,J., Helton,E., Kettner,M., Madan,A., Rodrigues,S.,			
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,			
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,			

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.,
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 2516)
 Strausberg, R.
 Direct Submission
 Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 On Aug 25, 2003 this sequence version replaced gi:18089244.
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue procurement: DCTD/BTP/Gazdar
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 14 Row: e Column: 22
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22749414.

FEATURES

source

1..2516
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:9042 IMAGE:3891543"
 /tissue_type="Lung, carcinoma, large cell undifferentiated."
 /clone_lib="NIH MGC_69"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1..2516
 /gene="ITM1"
 /note="synonyms: TMC, MGC9042"
 /db_xref="LocusID:3703"
 /db_xref="MIM:601134"
 93..2210
 /gene="ITM1"
 /codon_start=1
 /product="integral membrane protein 1"
 /protein_id="AAH20965.1"
 /db_xref="GI:18089245"
 /db_xref="LocusID:3703"
 /db_xref="MIM:601134"
 /translation="MTKFGFLRLSYEKQDTLLKLLLSMAVLFSFTRFLAVLRFSV
 IHEFDYFNFTTFLEAFEGFYKPHNFDRAWYPLGRIICGTIYPLGMLITGAAYVHV
 LHFPHITIDNVCVFLAPLPSFTTIVTHTLKLKADAGAGLAAALAVVPGYTSR
 SVAGSDNEGAIICMLTYMWTIKAVKGTSCIWAACALAYFWMYSGGVYFLNL
 IPLAVLMLTGRSHRIYVACTVYCLGILSNQISFQFQVLSSEHMAAFGVPGL
 CQIFAVDILSKNPQFVEFLRSVLISVGLVFLVLLVGLMLTGKISLFWTKRFFYSL
 DFTSAKNPIIASVSBEHQPTWSSYFDLIQVFPVGLYFCFNLSDARFIIMY
 GVTMYFSAVMRLMLVLPVCMILSGISGVLSVLYMKNLDSRPDKSKKQOOSTY
 PIKVEASGMILNMFALFLITPTFTSWTSEAYSSPSIVLSARGGSGSRILIFDDFRA
 YVLRHNTPEDAKVMWMDYQITAVANTILVDNNTWNTHTISRVGAMASTESKA
 VEIMRELDVSVLVI FGLTGYSSDDINKFLMVRIGGSTDTGKHKENDYXTPTEGF
 RVDREGSVLANCLWYKMYRFGQVTEAKRPGPDRVFNALGNKDKELDLVLEAY
 TTEHVLVYIKVLDNRGLSRT"

gene

CDS

ORIGIN

Alignment Scores:

Pred. No.: 7.86e-308 Length: 2516
 Score: 3657.00 Matches: 698
 Percent Similarity: 99.15% Conservative: 1
 Best Local Similarity: 99.01% Mismatches: 6
 Query Match: 98.92% Indels: 0
 Ds: 9 Gaps: 0
 US-10-028-384-12 (1-705) x BC020965 (1-2516)
 QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeu 20
 Db ATGACTAAGTTTGGATTTTTCGATTTTTCGATGAGAGCAGACACACTTTTGAAGCTT 152
 QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaLeuArg 40
 Db CTCATTCTGCAATGGCTGCTGATTATCTCTTCACCTCGTCTGTTGCTGCTCCGAGA 212
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
 Db TTTGAAAGTGTATCCATGAGTTTATCGCTACTTTAATATATCGGACTACCAAGTTCTTG 272
 QY 61 AlagluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrProLeu 80
 Db GCTGAGGAGGGGTTTATAAATCCATACTGTTTGTATGACCGAGCCCTGGTACCTTTTG 332
 QY 81 GlyArgIleleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db GGACGAATCATTTGAGGAAACAATTTACCAGGTTTAATGATCACCTCTGCTGCAATCTAC 392
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db CATGTACTCCCAATTTTCCACATCACCATCGAATTCGGAATGCTGTGTGTTCTCGGCC 452
 QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 Db CCTCTCTTCTCCTCTTACCACCATCGTCACTACCATCTTACCAAGAGCTCAAGGAT 512
 QY 141 AlaglyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
 Db GCAGGGGCTGGCTCTCTGCTGCCATGATTCGTAGTCTGCTGATATATCTCCCGA 572
 QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 Db TCTGTGGCTGGCTCCTATGATAATGAAGGAGTGGCCATCTTTTGCATGCTACTCACCTAC 632
 QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaLysCysAlaLeu 200
 Db TACATGTGGATCAAGGCAGTAAAGACTGGTTCCATCTGTGGGCAGCTAAGTGTGCCCTT 692
 QY 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIlePro 220
 Db GCTTATTTCTACATGGTCTGCTCATGGGAGGTTATGTGTTCTCTGATCAACTTAATTCCT 752
 QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 Db CTCACGCTCCCTGCTGCTGATGCTCACAGGCCGTTCTCTCACCGGATCATATGGGCCCTAC 812
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db TGTACTGTTTACTGCTGGGCACATATATCTTCTATGTCAGATCTCTTTTGGGGTTCCTAG 872
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheCysGlnIleHis 280
 Db CCTGTCTTCTCATCAGAGCACATGGCGCTTTTGGGGTCTTTTGGTCTCTGCCAGATCCAT 932
 QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db GCCTTTGTGATTACCTGGCAGCAAGTTGAATCCACAAATTTGAAGTTCCTTTTCCGG 992
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuMetLeuThr 320
 Db AGCGTCATCTCTCGTAGGCTTTGCTCTTCTCACCGTGGGAGCTCTCTCTCATGCTGAGA 1052

QY 321 GlyLysIleSerProThrThrGlyArgPheTyrSerLeuLeuAspSerTyrAlaLys 340
 DB 1053 GGAATAATATCTCCCTGGAGCGGGCTTTCTACTCGCTCGTGAATCCCTTTATGCTAAG 1112
 QY 341 AsnAsnIleProIleLeuAlaSerValSerGluHisGlnProThrThrTrpSerTyr 360
 DB 1113 AACAAATCCCATCATTTCTGTGTCTGAGCATCAGCCACAACTGTCCTCATAC 1172
 QY 361 TyrPheArgLeuGlnLeuValPheMetPheProValGlyLeuTyrCysPheSer 380
 DB 1173 TATTTTGACCTGACGCTCCCTGCTTCATGTTTCCAGTTGGCTCTATTACTGCTTTAGC 1232
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1233 AACCTGTCTGATGCCCGGATTTTATCATCATGTATGTGTGACCATGATGACTTTTCA 1292
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1293 GCTGTATGTGTGCTTAATGCTAGTGTGGACCTGTTATGTGCACTCTCTGGCAAT 1352
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 DB 1353 GGAGTCTCCAGGCTGCTGCATACATACATGAAGATCTGGACATAAGTCGTCCAGACAAG 1412
 QY 441 LysSerLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 DB 1413 AAGAGCAAGAGCAACAGGATTCACCTACCTATTAAAGTAAGTGGCAAGTGGATG 1472
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 DB 1473 ATACTGTGTCATGGCTTCTCTCTCATCACCTACCTCACTCACTCACTGGTGACCACT 1532
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 DB 1532 GAGGCTTACTCTCTCCGTCATGTACTATCTGCCCCGTGGGGATGGCAAGTAGGATC 1592
 QY 501 IlePheAspSerPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 DB 1593 ATATTTGATGACITCCGAGAAGCATATATTGCTTCGTCAATAATACTCCAGAGGATGCG 1652
 QY 521 LysValMetSerTrpTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 DB 1653 AAGTCTATGCTGGTGGATTTAGCTATACAGTTACAGCTATGCGAAACCAACAAAT 1712
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1713 TTAGTGACAATAACACATGGAATAATACCATATTTCTCGAGTAGGGCAGGCAATGGCG 1772
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1773 TCCACAGAGGAAAGGCTATGAGATCATGAGGAGCTCGATGTGAGTATGTGTGCTGTC 1832
 QY 581 IlePheGlyLeuThrThrTyrSerSerAspIleAsnLysPheIleThrTrpMetVal 600
 DB 1832 ATTTTGGAGGCTCTACTGGGTATCTCTGTATGATATCAACAAGTTCTTTGGATGTC 1892
 QY 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 DB 1892 CGGATTTGGAGGAGCACAGATACAGGCAACATATCAAGAGAAATGACTATTTACTTCCA 1952
 QY 621 ThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeuMetTyr 640
 DB 1952 ACTGGGAGTTCGGTGTGGACCGTGAAGTTCTCCAGTGTCTGCTCAACTGCTCATGTAC 2012
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 DB 2012 AAGATGTGTACTATCGCTTTGACAGGTTTACACAGAGGCCAGGCTCTCCAGGCTTT 2072
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 DB 2072 GACCGTCTCCGAATGCTGAGATGGGAATAAGACTTTGAGCTTGATGTCTCTGGAGAA 2132
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700

DB 2133 GCATATACACAGACATTTGGTGTGAGGATATACAGGTAAGGACCTGGATATCGA 2192
 QY 701 GlyLeuSerArgThr 705
 DB 2193 GCGTTGTCAAGGACA 2207
 RESULT 8
 LOCUS CQ272867 2472 bp DNA linear PAT 03-FEB-2004
 DEFINITION Sequence 13801 from Patent WO02068579.
 ACCESSION CQ272867
 VERSION CQ272867.1 GI:42295093
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, B.W.
 TITLE Kits, such as nucleic acid arrays, comprising a majority of
 humanexons or transcripts, for detecting expression and other uses
 thereof
 JOURNAL Patent: WO 02068579-A 13801 06-SEP-2002;
 PE Corporation (NY) (US)
 FEATURES
 source
 1..2472
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,09e-307 Length: 2472
 Score: 3652.00 Matches: 697
 Percent Similarity: 99.15% Conservatives: 1
 Best Local Similarity: 99.01% Mismatches: 6
 Query Match: 98.78% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-12 (1-705) x CQ272867 (1-2472)
 QY 2 ThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrIleuLysLeuLeu 21
 DB 110 ACTAAGTTTGGATTTTTCGATTGTCCTATGAGAGGAGGACACACTTTTGAAGCTTCTC 169
 QY 22 IleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPhe 41
 DB 170 ATTCTGTCATGGCTGCTGTATTATCTCTCCACTCGTCTGTTGCTGCTGAGATTT 229
 QY 42 GluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAla 61
 DB 230 GAAAGTGTATCCAGAGTTTGTATCCCGTACCTTAAATATCGACTACAGGCTCTCGCT 289
 QY 62 GluGluGlyPheTyrLysPheHisIleAsnTrpPheAspArgAlaTyrTyrProLeuGly 81
 DB 290 GAGGAGGGGTTTATAAATTCATTAACCTGTTGATGACCGAGCTGCTACCTTTGGGA 349
 QY 82 ArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyrHis 101
 DB 350 CGAATCATTTGGAGGAACAAATTTACCAGTTTAAATGATCACTCTGCTGCAATCTACCAT 409
 QY 102 ValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaPro 121
 DB 410 GTACTCCATTTTTCACATCACCATCGACATTCGAATGCTGTGTGTTCTTGGCCCT 469
 QY 122 LeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAspAla 141
 DB 470 CTCCTCTCTCTCTCTCCACCCATCGTACGACCTTACCAAGAGCTCAAGATGCA 529
 QY 142 GlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArgSer 161
 DB 530 GGGGCTGGGCTTCTTGTCTGCCCATGATGCTGTAGTTCCTGATATATATCTCCCATCT 589

QY 162 ValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyr 181
DB 590 GTGGCTGGCTCTATGATTAATGAAGAGGATGGCCATCTTTTGCATGCTACTACCTACTAC 649
QY 182 MetTrrPileLysAlaValIleThrGlySerIleCysTrrPAlaAlaLysCysAlaLeuAla 201
DB 650 ATGTGGATCAAGGAGTAAGACAGTTCCTCATCTGTTGGCAGCTAAGTGTGCCCTTGTCT 709
QY 202 TyrPheTyrMetValSerSerTrrPglyGlyTyrValPheLeuLeuLeuLeuProLeu 221
DB 710 TATTTCTACATGGTCTGCTGATGGGAGGTTATGTTCTGTGATCAACCTTAATTCCTCTC 769
QY 222 HisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaLysCys 241
DB 770 CACGTCCTCGTGTGATGCTCACAGGCGCTTCTCTCACCGATCTATGTCGCTACTGT 829
QY 242 ThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGlnPro 261
DB 830 ACTGTTTACTGCTGGGCACTATATCTTCTATGAGATCTCCTTTGGGTTTCCAGCCT 889
QY 262 ValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHisAla 281
DB 890 GTCCCTTTCATCAGAGCACATGGCAGCCTTTGGGCTCTTGGTCTCTGCCAGATCCATGCC 949
QY 282 PheValAspTyrLeuArgSerIleLeuAsnProGlnGlnPheGlnValLeuPheArgSer 301
DB 950 TTGTGGATTACCTCGCAGCAAGTTGAATCCACAATTTGAAGTTCTTTTCCGGAGC 1009
QY 302 ValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThrGly 321
DB 1010 GTCATCTCTGTTAGGCTTGTCTCTCACCGTGGGAGCTCTCTCTCATGCTGACAGGA 1069
QY 322 LysIleSerProTrrPThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLysAsn 341
DB 1070 AAAATATCTCTCGTGGGCGGCTTCTACTCGTGTGATCCCTCTTATGCTAGAGAAC 1129
QY 342 AsnIleProIleAlaSerValSerGluHisGlnProThrTrrPrrSerSerTyrTyr 361
DB 1130 AACATCCCATCATTTCTGCTGTCTGAGCATCAGCCCAACCTGCTCTCATACTAT 1189
QY 362 PheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSerAsn 381
DB 1190 TTGTACCTCGAGCTCCTCGTCTTATGTTCCAGTTGGCTCTATTAATCTCTTTAGAAC 1249
QY 382 LeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSerAla 401
DB 1250 CTGTCTGATGCCCGATTTTATCATCATCATGATGTTGTGACCAAGATGTTCTTTCAGCT 1309
QY 402 ValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIleGly 421
DB 1310 GTAATGGTGGCTTAATGCTAGTGTGGCACCTGTTATGTGCACTCTCTCTGGCATTTGA 1369
QY 422 ValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLysLys 441
DB 1370 GTCTCCAGGTCTGTCCACATACATGAAGAATCTGGACATAAGTCTCCAGACAAGAAG 1429
QY 442 SerLysLysGlnAlaAspSerThrTyrProIleLysIleGluValAlaSerGlyMetIle 461
DB 1430 AGCAAGAGCAACAGGATCCACCTTACCTTAAAGATGAAGTGGCAAGTGGATGATA 1489
QY 462 LeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTrrPrrValThrSerGlu 481
DB 1490 CTGGTCATGGCTTTCTTCTCATCATCACTTACCTTTCATCACTGGGTGACCATGTAG 1549
QY 482 AlaTyrSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIle 501
DB 1550 GCTACTCTTCTCCGTCATTTGATCTATCTCCCGGTGGTGGGATGGCAGTAGATCATA 1609
QY 502 PheAspAspPheArgGluAlaTyrTrrPrrLeuArgHisAsnThrProGluAspAlaLys 521
DB 1610 TTGTATGACTTCCGAGAGGACATATTATGGCTTCTGTCTATATACTCCAGGAGTGGAG 1669

QY 522 ValMetSerTrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrr 541
DB 1670 GTCATGCTCGTGGGATTAATGCTATCAGATTACAGCTATGGAACCAACCAATTTTA 1729
QY 542 ValAspAsnAsnThrTrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrr 561
DB 1730 GTGGACAATAACACATGGAATAATACCATATTTCTCGAGTAGGAGCAATGGCGTCC 1789
QY 562 ThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIle 581
DB 1790 ACAGAGGAAGAAGCCTATGAGTCACTGAGGAGCTGATGTACCTATGTGTGTCTCAT 1849
QY 582 PheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrrPrrMetValArg 601
DB 1850 TTTGAGGCGCTCACTGGGTATCTCTGATGATATCAACAAGTTCTTTGGATGGTCCGG 1909
QY 602 IleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTrrTrrPrrProThr 621
DB 1910 ATTGAGGGAGCACAGATACAGGCAACATATCAGAGAGATGACTATTATCTCAACT 1969
QY 622 GlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTrrLys 641
DB 1970 GGGGAGTTCCGTGTGGACCGTGAAGGTTCTCCAGTCTGCTCAACTGCTCATGTACAA 2029
QY 642 MetCysTrrTrrTrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrrPrr 661
DB 2030 ATGTTTACTATCTCTTGGACAGGTTTACAGAGCCAGGCTCTCTCAGGCTTTGAC 2089
QY 662 ArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGly 681
DB 2090 CGTGTCCGAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATCTCTGGAGGAGCA 2149
QY 682 TyrThrThrGluHisTrrPrrLeuValArgIleTrrLysValLysAspLeuAspAsnArgGly 701
DB 2150 TATACCACAGACATTTGGCTGTGTCAGATATACAGTAAGACCTGGATATTCAGGC 2209
QY 702 LeuSerArgThr 705
DB 2210 TTGTCAAGGACA 2221
RESULT 9
LOCUS HUMTMC 2293 bp mRNA linear PRI 13-FEB-2002
DEFINITION Homo sapiens transmembrane protein (TMC) mRNA, complete cds.
ACCESSION L47337.1
VERSION L47337.1 GI:18654193
KEYWORDS transmembrane protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2293)
AUTHORS Bellacosa, A.A.B. and Testa, J.J.R.
TITLE Isolation, characterization and mapping to human chromosome 11q24-25 of a gene, TMC, encoding a highly conserved putative transmembrane protein
JOURNAL Unpublished
FEATURES
Location/Qualifiers
1..2293
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="11q24-25"
/cell_type="mesothelioma"
/clone_lib="ddc11"
1..2293
/gene="TMC"
117..2234
/gene="TMC"
/note="putative"
/codon_start=1
/product="transmembrane protein"
/protein_id="AAL77539.1"

241	Qy	CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln	260
837	Db	TGTACTGTTTACTGCTGGGCACTATACTTCTATGCAGATCTCTTTGTGGGTTTCCAG	896
261	Qy	ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis	280
897	Db	CTGTGCTTTTCATCAGAGCAGATGCAGCCCTTTGGGGTCTTTGGTCTCTGCCAGATCCAT	956
281	Qy	AlaPheValAspTyrLeuArgSerIysLeuAsnProGlnGlnPheGluValLeuPheArg	300
957	Db	GCCTTTGGGATTACCTCGGAGCAGAGTTGAATCCACAACTTTGAAGTCTCTTTCCGG	1016
301	Qy	SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr	320
1017	Db	AGCGTCATCTCTCTGTTAGGCTTTGTCTTCTCACGTTGGAGCTCTCTCATGTGTGACA	1076
321	Qy	GlyIysIleSerProTrrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys	340
1077	Db	GGAAAAATATCTCCTGGACGGGGGTTCCTACTCGCTCTGGATCCCTCTTATGCTAAG	1136
341	Qy	AsnAsnIleProIleAlaSerValSerGluHisGlnProThrTrrpSerSerTyr	360
1137	Db	ACAACAATCCCATCATTTGCTTCTGTCTGAGCATCAGCCACAACTGGTCTCCATAC	1196
361	Qy	TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer	380
1197	Db	TATTTTGCCTCAGCTCCTCGCTTCATGTTTCCAGTTGGGCTCTATTACTGCTTTAGC	1256
381	Qy	AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer	400
1257	Db	AACTGTCTGATGCCGATTTTATCATCATGATATGGTGTACCAGCATGTACTTTTCA	1316
401	Qy	AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle	420
1317	Db	GCTGTAATGGTCGGTCTAATGTAGTTGGACCTGTTTATGTGCATTTCTCTGGCAIT	1376
421	Qy	GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys	440
1377	Db	GGAGTCTCCAGGTGCTGTCCATACATCATGAAGATCTGGACATAAGTCCCCAGACAAG	1436
441	Qy	LysSerLysLeuGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet	460
1437	Db	AAGAGCAAGAACACAGGATTCACCTACCTCTAATTAAGAATGAAGTGGCAAGTGGATG	1496
461	Qy	IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrrpValThrSer	480
1497	Db	ATACTGGTCATGGCTTCTTCTTCATCACTACCTACCTTTCATTCACCTGGGTGACCAGT	1556
481	Qy	GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyIleAspGlySerArgIle	500
1557	Db	GAGCGCTACTCTTCTCGTCCATTGTACTATCTGCCGTGATGGGATGGCAGTAGGATC	1616
501	Qy	IlePheAspAspPheArgGluAlaTyrTrrpLeuArgHisAsnThrProGluAspAla	520
1617	Db	ATATTTTGATGACTTCGAGAGGAGATATATTGGCTTCGTCATTAATTAATCTCCAGAGGATCG	1676
521	Qy	LysValMetSerTrrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle	540
1677	Db	AAGGTCAATGCTTGGTGGGATTATGGCTATCAGATTAACAGTATGGCAACCAAAAT	1736
541	Qy	LeuValAspAsnAsnThrTrrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla	560
1737	Db	TTAGTGACCAATTAACATAGAAATAATACCATATTTCTCGAGTAGGGCAGGCATGGCG	1796
561	Qy	SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal	580
1797	Db	TCCACAGAGAAAAGCCTATGAGATCATGAGGAGGCTCATGTTCAGCTATGTCTGGTCT	1856
581	Qy	IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrrpMetVal	600
1857	Db	ATTTTTGGAGCGCTCACTGGGTATCTCTGATGATATCAACAAGTTCTCGGATGGTC	1916

```

QY      601  A r g i l e G l y S e r T h r A s p T h r G l y L y s H i s I l e L y s G l u A s n A s p T y r T y r T h r P r o   620
Db      1917  C G A T T G G A G G A G C A C A G A T A C A G C A A C A T A C A A G G A A G A G A C T A T A T A C T C C A   1976
QY      621  T h r G l y G l u P h e A r g V a l A s p A r g G l u G l y S e r P r o V a l L e u L e u A s n C y s L e u M e t T y r   640
Db      1977  A C T G G G A G T T C C G T G T G G A C C G T G A A G T T C T C C A G T G C T C A A C T G C T C A T G T A C   2036
QY      641  L y s M e t C y s T y r T y r A r g P h e G l y G l n V a l T y r T h r G l u A l a L y s A r g P r o P r o G l y P h e   660
Db      2037  A G A T G T T A C T A C G T T T G G A C A G G T T T A C A G A A G C A A G C G C C T C C A G G C T T T   2096
QY      661  A s p A r g V a l A r g A s n A l a G l u I l e G l y A s n L y s A s p P h e G l u L e u A s p V a l L e u G l u G l u   680
Db      2097  G A C C G T G T C C G A A T G C T G A G A T T G G A A T A A A G A C T T T G A C T T G A T G T C T C G A G A A   2156
QY      681  G l y T y r T h r T h r G l u H i s T r p L e u V a l A r g I l e T y r L y s V a l L y s A s p L e u A s n A r g   700
Db      2157  G C A T A T A C A C A G A A C A T T G C T G T C A G G A T A T A C A A G S T A A A G G A C C T G G A T A A T C G A   2216
QY      701  G l y L e u S e r A r g T h r   705
Db      2217  G G C T T G T C A A G G A C A   2231

RESULT 10
HSM809254
LOCUS      HSM809254      2898 bp      mRNA      linear      PRI 30-AUG-2003
DEFINITION Homo sapiens mRNA; cDNA DKFZp79N2434 (from clone DKFZp79N2434).
ACCESSION   BX649102
VERSION     BX649102.1 GI:34368274
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2898)
AUTHORS     Lauber, J., Bahr, A., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
            Fobo, G., Han, M. and Wiemann, S.
CONSTRM     The German Human cDNA Consortium
TITLE       Direct Submission
JOURNAL     Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherrberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp79N2434) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES
Source      1..2898
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="DKFZp79N2434"
            /tissue_type="liver"
            /clone_lib="779 (synonym: hnccl). Vector pSport1_Sfi; host
            DH10B; sites SfiIA + SfiIB"
            /dev_stage="fetal"
            2817..2822
            polyA_signal
            polyA_site
            ORIGIN
            2863

Alignment Scores:
Pred. No.:      1,29e-306      Length:      2898
Score:          3644.00      Matches:      695
Percent Similarity: 98.87%      Conservatives: 2
Best Local Similarity: 98.58%      Mismatches: 8
Query Match:     98.57%      Indels:      0
DB:              Gaps:      0

```

```

US-10-028-384-12 (1-705) x HSM809254 (1-2898)
QY      1   M e t T h r L y s P h e G l y P h e L e u A r g L e u S e r T y r G l u L y s G l n A s p T h r L e u L e u L y s L e u   20
Db      186  A T G A C T A A G T T T G A T T T T T G C A T T G T C T A T A G A A G A G A C A C A C T T T T G A A G C T T   245
QY      21   L e u L e u S e r M e t A l a L a V a l L e u S e r P h e S e r T h r A r g L e u P h e A l a V a l L e u A r g   40
Db      246  C T A A T T C T G T C A A T G C T G C T G T A T T A T C C T T C C A C T C G T G T T T G T G T C T G C C T G A   305
QY      41   P h e G l u S e r V a l L e u H i s G l u P h e A s p P r o T y r P h e A s n T y r A r g T h r T h r A r g P h e L e u   60
Db      306  T T T G A A A G T G T T A T C C A T G A G T T T G A T C C G T A C T T T A A T T A T C G G A C T A C C A G G T T C C G   365
QY      61   A l a G l u G l y G l y P h e T y r L y s P h e H i s A s n T r p P h e A s p A r g A l a T r p T y r P r o L e u   80
Db      366  G C T G A G A G G G G T T T A T A A A T C C A T A A C T G G T T G A T G A C C G A G C C T G G T A C C C T T T G   425
QY      81   G l y A r g I l e I l e G l y L e u T h r L e u P r o G l y L e u M e t I l e T h r S e r A l a L a l e T y r   100
Db      426  G G A C G A A T C A T T G G A G A A C A A T T T A C C A G G T T T A A T G A T C A C C T C T G C T C A A T C T A C   485
QY      101  H i s V a l L e u H i s P h e P h e H i s I l e T h r I l e A s p I l e A r g A s n V a l C y s V a l P h e L e u A l a   120
Db      486  C A T G T A C T C C A T T T T T C C A C A T C A C C A T C G A C A T T C G G A T G C C T G T G T T C C T G G C C   545
QY      121  P r o L e u P h e S e r S e r P h e T h r S e r L e u V a l T y r T h r L e u L e u T h r L y s G l u L e u L y s A s p   140
Db      546  C C T C T C T C T C C T C C T T C A C C A C C A T C G T C A C G T A C C G C C T T A C C A A A G A G C T C A A G G A T   605
QY      141  A l a G l y A l a G l y L e u L e u A l a A l a M e t I l e A l a V a l P r o G l y T y r L e u S e r A r g   160
Db      606  G C A G G G G C T G G G C T T C T T G C T G C C A T G A T T G C T G A T T T C C T G G A T A T A T C T C C C G A   665
QY      161  S e r V a l A l a G l y S e r T y r A s p A s n G l u G l y L e u A l a l e P h e C y s M e t L e u L e u T h r T y r   180
Db      666  T C T G T G G T G G C T C C T A T G A T A A T G A G G G A T T G C A C T T T T T G C A T G C T A C T C A C T A C   725
QY      181  T y r M e t T r p I l e L y s A l a V a l L y s P h r G l y S e r I l e C y s T r p A l a A l a L y s C y s A l a L e u   200
Db      726  T A C A T G T G G A T C A A G G C A G T A A A G A C T G T T C C A T C T G T G G G A C G T A A G T G T G C C C T T   785
QY      201  A l a T y r P h e T y r M e t V a l S e r S e r T r p G l y T y r V a l P h e L e u I l e A s n L e u I l e P r o   220
Db      786  G C T T A T T T C T A C A T G C T C G T C A T G G G A G G T A T G T G T T C C T G A T C A C T T A A T T C C T   845
QY      221  L e u H i s V a l L e u V a l L e u M e t L e u T h r C l y A r g P h e S e r H i s A r g I l e T y r V a l A l a T y r   240
Db      846  C T C C A G C T C C T G C T G C T A T G C T C A C A G G C C G T T C T C T C A C C G G A T C T A T G T G C C C T A C   905
QY      241  C y s T h r V a l T y r C y s L e u G l y T h r I l e L e u S e r A r g G l n I l e S e r P h e V a l G l y P h e G l n   260
Db      906  T G T A C T G T T A C T G C C T G G G C A C T A T A C T T T C T A T G C A G A T C T C T T T G T G G G T T T C A G   965
QY      261  P r o V a l L e u S e r S e r G l u H i s M e t A l a G l y P h e G l y V a l P h e G l y L e u C y s G l n I l e H i s   280
Db      966  C C T G T C C T T T C A T C A G A C A C A T G C G A C C C T T T G G G G T C T T T G G T C T C T G C A G A T C C A T   1025
QY      281  A l a P h e V a l A s p T y r L e u A r g S e r L y s L e u A s n P r o G l n P h e G l u V a l L e u P h e A r g   300
Db      1026  G C C T T T G T G A T T A C C T G C G C A G A G T T G A T C C A C A A C A A T T T G A A G T T C T T T C C G G   1085
QY      301  S e r V a l L e u S e r L e u V a l G l y P h e V a l L e u L e u T h r V a l G l y A l a L e u M e t L e u T h r   320
Db      1086  A G C G T C A T C T C T G T G A G G C T T T G C C T T C T A C C C T G G A G G C T C C T C A T G C T G C A C A   1145
QY      321  G l y L y s I l e S e r P r o T r p T h r G l y A r g P h e T y r S e r L e u A s p P r o S e r T y r A l a L y s   340
Db      1146  G G A A A A A T A T C C C T C G A C G G G G C G T C T A C T C G C T G C T G G A T C C C T C T T A T G T A A G   1205
QY      341  A s n A s n I l e P r o I l e L e u A l a S e r V a l S e r G l u H i s G l n P r o T h r T h r T r p S e r S e r T y r   360
Db      1206  A A C A A C A T C C C A T A T T G T T C T G T C T G A G C A T C A G C C C A C A C A C C T G C T C C T C A T A C   1265

```

QY 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 Db 1266 TATTTGACCTGCAGCTCCTCGCTTCATGTTTCCAGTTGGCTCTATTACTGCTTTAGC 1325
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 Db 1326 AACTGTGCTGAGCCCGGATTTTATCATCATGATATGGTGTACCCAGCACTGACTTTTCA 1385
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 Db 1386 GCTGTAATGGTGGCTTAATGCTAGTGTGGCACTGTTATGTGCATTCTCTCGCAT 1445
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLeuAsnLeuAspIleSerArgProAspIys 440
 Db 1446 GGATGCTCCAGGGTGTGTCCACATCATGAAGAATCTGGACATAGTGTCCAGACAG 1505
 QY 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 Db 1506 AAGAGCAAGAACCAACAGGATTCACCTACCTATTAAAGATGAAGTGGCAAGTGGATA 1565
 QY 461 IleLeuValMetAlaPheLeuIleThrTyrPheHisSerThrThrValThrSer 480
 Db 1566 ATACTGGTCATGGCTTCTTCTTCATCACCCTACCTACCTTCATCAACCTGGGTGACCA 1625
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 Db 1626 GAGGCTACTCCTCTCCGTCATTTACTATCTGCGCTGGTGGGATGGCAGTAGGATC 1685
 QY 501 IlePheAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProGluAspAla 520
 Db 1686 ATATTGATGACTCCAGAGCAATATTATGCTTCGTCATTAATCTCCAGAGATGG 1745
 QY 521 LysValMetSerTyrPheTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 Db 1746 AAGGTATGCTCCTGGTGGATATGGCTATCAGATTACAGCTATGGCAACCAACAT 1805
 QY 541 LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaValAla 560
 Db 1806 TTAGTGACATAACATGAGATATATCCCATTTCTCGAGTAGGCGGCAATGGCG 1865
 QY 561 SerThrGluGlyLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1866 TCCACAGAGGAAAAAGCCTATGATCATGAGGAGCTTGATGTACGCTATGTGCTGTC 1925
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuThrPheVal 600
 Db 1926 ATTTTGGAGCCCTCACTGGGTATCTCTGATGATATCAACAGTTCTTTGGATGGTC 1985
 QY 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1986 CGGATGGAGGAGCACAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCA 2045
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 2046 ACTGGGAGTTCCGTGTGGACCGTGAAGGTTCCTCAGTGTCTCAACTGCCTCATGTAC 2105
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 2106 AAGATGTGTACTATCGCTTGGACAGGTTTACACAGAGCCAGCGCTCTCCAGGCTTT 2165
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 2166 GACCGTGTCCGAATGCTGAGATGGGAATTAAGACTTTGAGCTTGATGCTCTGGAGAA 2225
 QY 681 GlyTyrThrGluHisThrPheValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2226 GCATATACCAAGAACATGGCTGCTCAGGATATACAGGTAAGGACCTCGGATATCGA 2285
 QY 701 GlyLeuSerArgThr 705
 Db 2286 GCCTGTGTACAGGACA 2300

RESULT 11

CQ834710

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

CQ834710
 Sequence 581 from Patent WO2004058805.
 CQ834710.1 GI:50834247
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 1. Matsuda A. and Yoneta S.
 T cell activating Gene
 Patent: WO.2004058805-A 581 15-JUL-2004;
 Asahi Kasei Pharma Corporation (JP)
 Location/Qualifiers
 1..2435
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 73..2187
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAH05570.1"
 /translation="MTKFLRLSYEKQDTLLKLLSMAAVLSFSLFAVLRFESV
 IHEFDYFNRYRTRELAEGYKFNHFDRAWYPLGRIIGGTYPLGMIATSAIYHV
 LHPFTTIDRNVCVFLAPLSPFTTIVYHLTKELDAGAGLAAAMIAVVPQSR
 SVAGSDNEGAIKPCMLTYVMKAVKATGSIKAAKALAFVWVSSMGVYVFLNLI
 QHAFVLMILTSPFPHRIYVAYCTVYCLGTILSMQISFVGPVLSSHMAAFGVGLC
 PLHAFVYHRSKLNQFQFVLRFSVLSUGFVLITVGLLMTKISFWTGRYSLLD
 PSYANNPIIASVSEHPTWSSYFDLQLLVFMFVGLYFCNSLSDAKIFLITMG
 VTSMTFSAVMRLMLVAPVNCILSGISQVLSTYKKNLDISRPDKSKKQDSTYP
 IKNEVASGMILVWAFLLTYTFTSTWTSSEAYSPSIVLSARGDGSRIIFDFREAY
 YWLKNTPEDAKMSMDYGYQITAMANRTILVDNNTWNNTHSRVQQAASATEKAY
 EIMRELDVSYLVIFEGTLGYSDDKINKFLMWIRIGSTDTGTHIKENDVYTPGEPR
 VDREGSVLLNCLAYKMCYRFGQVYTEAKRPPGPDVRVNAEIGNKDFELDLVEAYT
 TEHLVRIYKVKLDLNRGLSRT"

ORIGIN

Alignment Scores:

Pred. No.: 8,29e-306 Length: 2435
 Score: 3633.50 Matches: 696
 Percent Similarity: 98.87% Conservative: 1
 Best Local Similarity: 98.72% Mismatches: 7
 Query Match: 98.28% Indels: 1
 DB: 6 Gaps: 1

US-10-028-384-12 (1-705) x CQ834710 (1-2435)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 Db 73 ATGACTAAGTTTGGATTTTGGATTTGCTTATGAGAGCAGGACACACTTTTGAAGTT 132
 QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 Db 133 CTCATTCTGTCAATGGCTGCTCTATTATTCCTTCTCCACTCGCTGTGTGCTGCTGAGA 192
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 Db 193 TTTGAAGTGTTATCCATGAGTTTATCGACTTTTATTTATTCGAGCTACCCAGGTCTCTG 252
 QY 61 AlaGluGluGlyPheTyrLysPheHisAsnThrPheAspAspArgAlaTyrTyrProLeu 80
 Db 253 GCTGAGAGGGGTTTATTAATATCCATACTGGTTGATGACCGAGCTGTGTACCTTTG 312
 QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 313 GGACGAATCATTCGAGGAGCAATTTACCAGGTTTATATGATCACCTCTGCTGCAATCTAC 372
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120

Db 373 CATGTACTCCATTTTTCACATACACCATCGACATTCGGAATGCTGTGTGTTCTCTGGCC 432
 Qy 121 ProLeuPheSerSerPheThrSerIleValThrThrLeuLeuThrLysGluLeuLysAsp 140
 Db 433 CCT 492
 Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
 Db 493 GCAGGGGCTGGGCTTCTCTGCTGCTGCCATGATGCTGTAGTTCCTGGATATATCTCCCGA 552
 Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 Db 553 TCTGTGGCTGGCTCTCTATGATGATGAGGAGTTCGCCATCTTTTGCATGCTACTCACCTAC 612
 Qy 181 TyrMetTyrIleLeuAlaValLysThrGlySerIleCysTyrIleAlaAlaLysCysAlaLeu 200
 Db 613 TACATGTGCATCAAGCAGTAAGACTGGTTCATCTCTGTTGGGAGCTAAGTGTGGCCCTT 672
 Qy 201 AlaTyrPheTyrMetValSerSerTyrPheGlyTyrValPheLeuIleAsnLeuIlePro 220
 Db 673 GCT---TTCATCATGGTCTCTGTCATGGGAGGTTATGTTCTCTGATCTCACTTAATCTCT 729
 Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 Db 730 CTCCAGCTCCTCTGCTGATGCTCACAGGCGGCTTCTCTCACCGGATCTATGTGGCCTAC 789
 Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db 790 TGTACTGTTTACTGCTGGGCACTATATCTTCTATGCAAGTCTCTTGTGGGTTCAG 849
 Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 Db 850 CCTGTCTTCTATCAGAGCACATGGCAGCCTTTGGGGTCTTTGGTCTCTCTGCCAGATCCAT 909
 Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db 910 GCCTTTGTGGATTACTGGCGAGCAAGTTGATCCACACAAATTTGAGTCTCTTTCCGG 969
 Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 Db 970 AGCGTCATCTCTCTGTAGGCTTTCTCTCTCTCACCGTGGGAGCTCTCTCATGCTGACA 1029
 Qy 321 GlyValIleSerProTyrPheThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 Db 1030 GGAATAATATCTCCCTGGAGGGGGCTTTCTACTCGCTGCTGGATCCCTCTTATGCTAAG 1089
 Qy 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 Db 1090 AACAAATCCCATCATTCCTCTGTCTGCTGAGCATCAGCCCAACACCTGGCTCTCATAC 1149
 Qy 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 Db 1150 TATTTTGACCTGAGCTAGC 1209
 Qy 381 AsnLeuSerAspAlaArgIlePheIleLeuMetTyrGlyValThrSerMetTyrPheSer 400
 Db 1210 AACCTGCTGTGATGCCCGGATTTTATCATCATGATGATGTTGTGACCATGCTACTTTTCA 1269
 Qy 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 Db 1270 GCTGTAATGGTGGCTTAATGATGATGTTGGACCTCTTATGTCATCTCTCTCTGGCAT 1329
 Qy 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 Db 1330 GGAGTCTCCAGGTGTGTCACATACACAGAGATCTGGACATAAGTCGTCACAGACAAG 1389
 Qy 441 LysSerLysGlnGlnAspSerThrTyrProIleValIleGluValAlaSerGlyMet 460
 Db 1390 AAGAGCAAGAGACACAGATTTCCACTTACCTATTAAGATGAAGTGGCAAGTGGGATG 1449
 Qy 461 IleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThrSer 480
 Db 1450 ATACTGGTCATGGCTTCTCTCTCATCTACCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1509

Qy 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 Db 1510 GAGGCTTACTTCTCTCGTCCATTTGTAATCTCTCCCGTGTGGGATGTCAGTAGCATC 1569
 Qy 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAlaAla 520
 Db 1570 ATATTGTGATGACTTCCGAGAGCATATATTGTGCTTCTCATATACTCTCAGAGGATGCG 1629
 Qy 521 LysValMetSerTyrTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 Db 1630 AAGGTCAATGCTCTGTGGGATTAATGCTATCAGATTACAGCTATGCAACCAACAATT 1689
 Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1690 TTAGTGGACAATAACATACATGGAATATACCCATATTTCTCGAGTAGGAGCAATGGCG 1749
 Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1750 TCCACAGAGGAAAAAGCCTATGAGATCATGAGGAGGCTCGATGTCAGCTATGTGCTGTC 1809
 Qy 581 IlePheGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 Db 1810 ATTTTGGAGGCTCTCACTGGGATTTCTCTGATGATATCAACAGATTCTTTGATGGTGC 1869
 Qy 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1870 CGGATTGGAGGAGCACAGATACAGGCAACATATCAAGGAGATGACTATTACTCCA 1929
 Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1930 ACTGGGAGTTCCTGTTGGACCGTGAAGGTTCTCCAGTGTCTCAACTCCCTCATGTAC 1989
 Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 1990 AAGATGTCTTACTATCTGCTTTGGACAGCTTTACACAGACCAAGCGCTCTCCAGGCTTT 2049
 Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
 Db 2050 GACCGTGTCCGAAATGCTGAGATTGGGAAATAAGACTTTTGGCTTGTGCTGGAGGAA 2109
 Qy 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2110 GCATATACACAGAACATTTGCTGTCAGGATATACAAAGTAAAGGACCTGGATATCGA 2169
 Qy 701 GlyLeuSerArgThr 705
 Db 2170 GGCCTTGTCAAGGACA 2184

RESULT 12

BC037612

LOCUS

DEFINITION

MGC:46919 IMAGE:535379), complete cds.

ACCESSION

BC037612

VERSION

BC037612.1

KEYWORDS

MGC.

SOURCE

Mus musculus

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 2170)

AUTHORS

Strausberg, R.B., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Schaeetz, P.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Carninci, P., Prange, C., Raha, S.S., Iqbal, N.A., Peters, G.J.,

Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

2736 bp mRNA linear ROD 30-JUN-2004

Mus musculus integral membrane protein 1, mRNA (cdna clone

MGC:46919 IMAGE:535379), complete cds.

BC037612

BC037612.1

MGC.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2170)

Strausberg, R.B., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Schaeetz, P.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Carninci, P., Prange, C., Raha, S.S., Iqbal, N.A., Peters, G.J.,

Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzinski, M.I., Skalek, U., Smal, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903. (2002)
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	2 (bases 1 to 2736)
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Strausberg, R.
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Submitted (13-SEP-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	NIH-MGC Project URL: http://mgc.nci.nih.gov
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Contact: MGC help desk
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Email: cgaps@mail.nih.gov
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Tissue Procurement: Jeffrey Green M.D.
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	cDNA Library Preparation: Life Technologies, Inc.
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Gaithersburg, Maryland
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Web site: http://www.nisc.nih.gov/
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Contact: nisc.mgc@nih.gov
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Akhtar, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Series: IRAC Plate: 81 Row: n Column: 21
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6680497.
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	Location/Qualifiers
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	1. 2736
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/organism="Mus musculus"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/mol_type="mRNA"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/strain="FVB/N"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/db_xref="taxon:10090"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/clone="MGC:46919 IMAGE:5353797"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal carcinoma, 5 month old virgin mouse."
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/clone_lib="NCI CGAP Mam6"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/lab_host="DH10B"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/note="vector: pCMV-SPORT6"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	1. 2736
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/gene="Itm1"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/db_xref="LocusID:16430"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/db_xref="MGI:105124"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	108. 2225
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/gene="Itm1"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/codon_start=1
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/product="integral membrane protein 1"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/protein_id="AAH37612.1"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/db_xref="GI:22902289"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/db_xref="LocusID:16430"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/db_xref="MGI:105124"
JOURNAL PUBMED REFERENCE AUTHORS JOURNAL	TITLE	/translation="MTKLGLRLSKQDRLKLLLSMAVLSFSTRFLRPFVSRV IHEPDPVNTKTRFLAEQGFYFHNFDRAWPLGRIGITVPLGLMITGAIIYHV LHPHIIIDIRNVCVFLAPLFSFTTIVTLTKELKADAGLAAAMIAVVPYISR SVAGSNEGIAIFCMLLTYVMKTKGSIYAAKCALAFYVYSSXGQVYFVLN IPLHVLVMTGRFSRIYAYCTVYCLGILSLSQISFVGFPVLSSEHMAAFVGL CQIHAFVDYLRSLNPOQFEVLFRSVLSLVGVLLTVGLLMTGKISFWTRFVSL

ORIGIN	
Alignment Scores:	
Pred. No.:	1,08e-305
Score:	3633.00
Percent Similarity:	98.87%
Best Local Similarity:	98.44%
Query Match:	98.27%
DB:	10
Gaps:	0
US-10-028-384-12 (1-705) x BC037612 (1-2736)	
Qy	1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db	108 ATGACTAAGCTTGGATTTTGGGATTTGCTCTATGAGAGCAGACACACTTCTTAAGTT 167
Qy	21 LeuLeuSerMetAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db	168 CTATCTCTGTCGATGGCTGCTGTTATCTTTTCTACTCGTCTTTTGTCTGCTGAGA 227
Qy	41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db	228 TTTGAAAGTGTATCATCCATGATTTGATCGGTACTTTTAAATATCGGACTACCGGTTCTG 287
Qy	61 AlaGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTyrProLeu 80
Db	288 GCTGAGAGGGGTTTATAAATCCATAACTGTTGATGACCGGCTTGTACCCCTTG 347
Qy	81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100
Db	348 GCGCAATCATTTGGAGGACAAATTTACCAGGTTTAAATGATCACTTCTGCTGCAATCTAC 407
Qy	101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db	408 CATGTACTCCATTTCTTCCATATCACTATTGACATTCGGAATGCTGTGTTTCTCGGCC 467
Qy	121 ProLeuPheSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db	468 CCATTTTCTTCCTTTCACCAACATTCGTTACCTACCAAGAGCTCAAGGAT 527
Qy	141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
Db	528 GCAGGAGCTGGGCTTCTTGTGCTGCCATGATTGCTGTAGTCTCTGGGTATATTTCTCGA 587
Qy	161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db	588 TCTGTAGTCTGCTCTCTATGATATGAGGAATTTGCTATCTTTTGCATGCTGCTTACTTAC 647
Qy	181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaLysCysAlaLeu 200
Db	648 TACATGTGGATCAAGGAGGAGTCAAGACTGGTTCCTATTTGGGCTGCAAGTGTGCCCTC 707
Qy	201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIleAsnLeuIlePro 220
Db	708 GCTTATTTCTACATGGTCTCTTTCATGGGAGGCTATGTGTTCTGATCAACTTGAATCTCT 767
Qy	221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db	768 CTACATGCTCTGGTGTCTAATGCTGACAGCGCGTTTTTCTCAGCGGATCTACGTAGCTAC 827
Qy	241 CysThrValTyrCysLeuGlyThrIleSerSerArgGlnIleSerPheValGlyPheGln 260
Db	828 TGTACTGTGTTACTGCTGGGACCACTTCTTCTATGAGATTTCTTGTGTTGTTTCCAG 887
Qy	261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280

Db 888 CCGTCCTTTTCATCAGAACACATGCGCTTTGGAGTGTGTTGGTCTCTGTCAGATCCAT 947
Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
Db 948 GCTTTTCGTAGATTACTCGCAGCAAGTTGAATCCACAGCAATTCGAAGTTCCTTTCCGG 1007
Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db 1008 AGTGTAATCTCCCTGGTTCCT 1067
Qy 321 GlyIleValSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db 1068 GGAATAATTTCTCCCTCGACAGGGCGTTTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1127
Qy 341 AsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTyrSerSerTyr 360
Db 1128 AATRAATTCATATTGTCATCTGTTCTGAGCACAGCCACCACTGGTCTCTCTCTCTCTCT 1187
Qy 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
Db 1188 TATTTTGATCTACAGCT 1247
Qy 381 AsnLeuSerAspAlaArgIlePheIleLeuMetTyrGlyValThrSerMetTyrPheSer 400
Db 1248 AACCTGCTGATGCTCGGATTTTATCATCATGATGATGATGATGATGATGATGATGATGAT 1307
Qy 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
Db 1308 GCTGTAATGTCGCTCTAATCTGCTGATTTGGACCTGTTATGTCATCTCTCTCTCTCTCT 1367
Qy 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
Db 1368 GGTGTTTCCAGGTCGTCACATATGAAATCTGACATACTGACATACTGACATACTGACATA 1427
Qy 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
Db 1428 AAGAGCAAGAACCAACAGGATCTTACCTTATTAAAGATGAGGTGGCGATGGGATG 1487
Qy 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTyrValThrSer 480
Db 1488 ATACTGTCATGGCTTTTCTCATCACCTACACGTTTCACTTCGATCTGGGTGACAGT 1547
Qy 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
Db 1548 GAAGCCTATTCTCTCCCTCCATGTAAGTCTGCTGCTGGGATGGCAGTAGGATC 1607
Qy 501 IlePheAspAspPheArgGluAlaTyrTyrThrLeuArgHisAsnThrProGluAspAla 520
Db 1608 ATTTTTCATGACTTCCGAGAGCGTATATTGGCTCCGTCCATCAATCTCCAGAGGATGCA 1667
Qy 521 LysValMetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
Db 1668 AAGTCATGTCATGGTGGATTTATGGCTACCAATTTACTGCAATGGCAATCGCAAT 1727
Qy 541 LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1728 TTAGTGACAAATAACACATGGAATAATACCATATTTCTGAGTAGCGGAGCAATGGCA 1787
Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
Db 1788 TCCACAGAGAAAGAGCTATGAATCATGAGGAGCTTGATGTCAGTATGCTGTGTC 1847
Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAlaAsnLysPheLeuTyrMetVal 600
Db 1848 ATTTTGGAGGCTTACTGGTATTCTTCGATGATATCAACAGTTCTTTGGATGGTC 1907
Qy 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
Db 1908 CGATTGGAGAGACACAGACAGGAGACACATTAAGAGAAATGACTACTATATCTCT 1967
Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1968 ACTGGGGAATTCGTGTTGATCGTGAGGGTTCCTCGGTCTGCTCACTGCTTATGTAC 2027

Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
Db 2028 AAAATGTGTACTACCGCTTTGGCAGGTCTACACAGAAGCAAGCTCCACAGGCTTT 2087
Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
Db 2088 GACCGTTCGAAATCTGAGATTGGTAATAAAGACTTTTGAAGCTTGATGCTCGAGGAA 2147
Qy 681 GlyTyrThrThrGluHisTyrLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
Db 2148 GCGTATACACAGAACACTGGCTAGTCAGATATACAGGTAAAGGACCTCGGTAATCGA 2207
Qy 701 GlyLeuSerArgThr 705
Db 2208 GGCTTGTCAAGGACA 2222
RESULT 13
AX799090
LOCUS AX799090 3094 bp mRNA linear PAT 10-OCT-2003
DEFINITION Sequence 9 from Patent WO03054008.
ACCESSION AX799090
VERSION AX799090.1 GI:37605062
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Perreault, C. and McBride, K.
TITLE Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy
JOURNAL Patent: WO 03054008-A 9 03-JUL-2003;
COMPAT: gene Inc. (CA)
FEATURES
Location/Qualifiers
1..3094
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/note="ITM1 accession No NM-008408"
ORIGIN
Alignment Scores:
Pred. No.: 1.27e-305 Length: 3094
Score: 3633.00 Matches: 694
Percent Similarity: 98.87% Conservative: 3
Best Local Similarity: 98.44% Mismatches: 8
Query Match: 98.27% Indels: 0
DB: 6 Gaps: 0
US-10-028-384-12 (1-705) x AX799090 (1-3094)
Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
Db 112 ATGACTAAGCTGGATTTTTCGATTGCTCTATGAGAGAGGACACACTTCTTAAGCTT 171
Qy 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 172 CTCATCTCTGTCGATGGCTGCTGTATCTTTTCTACTCTCTTTTCTGCTGCTGCTGAGA 231
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 232 TTTGAAAGTGTCATCCATGAGTTTATCGGCTTTTAAATATCGGACTACCCGGTTTCTG 291
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTyrTyrProLeu 80
Db 292 GCTGAGGAGGCTTTTATAATTCATACTGTTTTCATGACCGGCTTGTGACCTTTG 351
Qy 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 352 GCCCAATCATGTGAGGAAACAATTTACCCAGGTTTAAATGATCACTCTCTGCTCAATCTAC 411
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120


```

/strain="BALB/c"
/sub_species="domesticus"
/db_xref="taxon:10090"
/chromosome="9"
/tissue_type="mandibular condyle"
1.3094
/gene="Itm1"
1.111
/gene="Itm1"
112.2229
/gene="Itm1"
/notes="putative"
/codon_start=1
/product="integral membrane protein 1"
/protein_id="A047775.1"
/db_xref="GI:508543"
/translation="MYKLGFLRLSYEKQDTLLKLLLSMAAVLSFSTRFLFAVLRESV
LHEFDIFNIRTRFAESGFYKPHNWFDDRWYPIGRIGGTIYPLGMITSAAIYHV
LHFPHITDIRNVCVFLAPLFSFTTIVYHLTKELDKAGLLAAAMIAVPGYISR
SVAGSYDNEGIAIFCMLLTYMMIKAVKTSIYMAAKCALAFYVMVSSGGYVFLINL
IPLHVLAVLIGTRFSHRIYVAYCTVYCLGTILSMQISFVGFQPVLSSEHMAAFGVFL
COLHAFVDYLRSLNPOQEVLPFRSVISLGVFLVLTGALLMLTGKISPWTCGRFYSL
DPYAKNNIPILIASVSEHOPTWSSVYFDLQLLVFMPVGLYVCSNLSDAEFLIMY
GVTSMYSAWMLMLVLAFCILSGIGVSQVLSYMKNLDISRPDKSKKQDQSTY
PIKNEVASGMILWAFPLITYTPHSTWVTSEAYSSPSIVLSARGDGSRIIFDDPREA
YWLRENTPEDAKMSWDYGOITAMANRTILVDNNTWNTHISRVGQAMASTBEKA
YEIMRELDSYVLVIFGGLTGXSSDDINKFLMWRIIGGSTETGRHKENDYYTPTEGF
RVDRGSPVLLNCLMYKVCYRFGQVYTEAKRPPGDFRVNAEIGNKDFELDLSEAY
TTEHMLVRIYKVKDLNRLSRT"
112.201
/gene="Itm1"
202.2226
/gene="Itm1"
/product="transmembrane protein"
/notes="putative"
2230.3094
/gene="Itm1"
2443.2448
/gene="Itm1"
/notes="putative"
2470
/gene="Itm1"
/notes="putative"
2704.2709
/gene="Itm1"
/notes="putative"
2723.2728
/gene="Itm1"
/notes="putative"
2725
/gene="Itm1"
/notes="putative"
2738
/gene="Itm1"
/notes="putative"
3075.3080
/gene="Itm1"
/notes="putative"
3093
/gene="Itm1"
/notes="putative"

sig_peptide
mat_peptide
3'UTR
polyA_signal
polyA_site
polyA_signal
polyA_signal
polyA_site
polyA_site
polyA_signal
polyA_site
ORIGIN
Alignment Scores:
Pred. No.: 1-27e-305 Length: 3094
Score: 3633.00 Matches: 694
Percent Similarity: 98.87% Conservative: 3
Best Local Similarity: 98.44% Mismatches: 8
Query Match: 98.27% Indels: 0
DB: 10 Caps: 0

US-10-028-384-12 (1-705) x MUSTRRP (1-3094)

```

1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
112 ATGACTAAGCTTGGATTTTTCGGATTGTCTATGAGAAGCAGCAGACACTTCTAAGCTT 171
21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
172 CTCATCTCTGTCGATGGCTGCTGTGTATCTTTTCTACTCGTCTTTTGTGCTGAGA 231
41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
232 TTTGAAGTGTCAATCCATGAGTTTATCGGACGTTTAAATTATCGGACTACCGGTTCTG 291
61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaIleTyrProLeu 80
292 GCTGAGGAGGGGTTTATAAATTCATAAATCGGTTGTATGACCGGGCTTGGTACCTTTG 351
81 GlyArgIleLeuGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
352 GGCAGATCATTTGGAGGAACAATTTACCCAGGTTTAATGATGACCTTCTGCTGCAATCTAC 411
101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheAla 120
412 CATGTACTCCATTTCTTCCATATCATATTCACATTTCGAAATGCTGTGTTTTCTGGCC 471
121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
472 CCACCTTTCTCTCTTTTCAACCACCATCGTTACGTACCACTTTACCAAAGAGCTCAAGAT 531
141 AlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
532 GCAGAGCTGGGCTTCTTGTCTGCCATGATTCCTGTAGTTCCTGGGTATATTTCGGA 591
161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
592 TCTGTAGCTGCTCTCTATGATAATGAAGGAATTCCTATCTTTGCAATGCTGCTACTTAC 651
181 TyrMetThrIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
652 TACATGTGGATCAAGGAGCTGAAGACTGGTTCATCTATTGGGCTGCAAGTGTGCCCTC 711
201 AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuIlePro 220
712 GCTTATTTCATACATGGTCTCTTCATGGGAGGCTATGTGTCTCGATCAACTGATTCTCT 771
221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
772 CTACATGTCTCTGTGCTAAATGCTCAGAGGCGGTTTTTCTCACCGGATCTACGTAGCCTAC 831
241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
832 TGTACTGTCTACTGCTGGGACCATCTTCTATGACAGATTTCTCTTTGTTGTTTCCAG 891
261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
892 CCCGCTCTTTCATCAGAACACATGGCAGCCTTTGGAGTGTTTGCTCTCTGATGATCCAT 951
281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnInPheGluValLeuPheArg 300
952 GCTTTCGTAGATTACCTGCGCAGCAAGTTGAATCCACAGCAATTCGAAGTCTCTTTCCGG 1011
301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
1012 AGTGTATCTCCCTGGTGGCTTTGTCTCTCTCCTGCTGGAGCTCTCTCTCATGCTAACCA 1071
321 GlyLysIleSerProTrpThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
1072 GGAATAATTTCTCCCTGGACAGGCGGTTTCTACTCTCTGCTGGATCCCTCTTATGTAAG 1131
341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
1132 AATAACATTTCCATTATGATCTGTTTCTGAGCAGCCGCCAACCTGCTCTCTCTAC 1191
361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380

1192 TATTGTTGATCTACAGCTCCTTGTCTTCATGTTCCAGTTGGCCCTCTATTACTGCTTTAGC 1251
 381 AsnLeuSerAspAlaArgPheIleMetTyrGlyValThrSerMetTyrPheSer 400
 1252 AACCTGCTGATGCTCGGATTTTATCATCATGATGATGTTGACAGCATGACTTTTCA 1311
 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 1312 GCTGTAATGCTGCTGCTAATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 1371
 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 1372 GGTGTTTCCAGTGCTGCTCCATATATGAAATCTGGACATAGTCGCCAGAGAG 1431
 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 1432 AAGAGCAAGAGCAACAGCAATCTACTTACCCTATTAAAGATGAGGTGGCGAGTGGATG 1491
 461 IleLeuValMetAlaPhePheIleThrTyrPheHisSerThrTyrValThrSer 480
 1492 ATACTGGTTCATGGCTTTTCTCTATCACCACACGTTTCACTTGCATTCAGTTGGTACCAGT 1551
 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 1552 GAAGCCTATCTTCCCTCCATCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1611
 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 1612 ATTTTGTGATGACTTCCGAGACGGTATTATTGGCTCCGCTCAATACATCTCCAGAGGATGCA 1671
 521 LysValMetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 1672 AAGTCATGTCATGTTGGATTTATGGCTACCAATTAATCTGCAATGGCAATTCGACAAAT 1731
 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 1732 TTATGTGCAATAAACAATGGAATTAATACCAATTTCTCGAGTAGGCGCAATGGCA 1791
 561 SerThrGluGluValAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 1792 TCCACAGAGAAAGACCTATGAAATCATGAGGAGCTTGATGTCAGTATGCTGTGTC 1851
 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTyrMetVal 600
 1852 ATTTTGGAGGCTTACTGGGTATCTTCGGATGATATCAACAGCTTTCTTTGGATGGTC 1911
 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 1912 CGGATTGGAGGAGACAGACAGAGACAGACATTAAGAGATGATGACTACTTACTCTCT 1971
 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 1972 ACTGGGGAATTCGGTTCATCGTGGAGGTTCTCCGGTCTGCTCAACTGCTTATGTAC 2031
 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 2032 AAAATGTGTACTACCGCTTTGGGAGGTCATACAGAGCCACAGCGTCCACAGGCTTT 2091
 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 2092 GACCGTTCGAAATGCTCAGATTGGTAATAAGACTTTGAGCTTGATGCTCTGGAGAA 2151
 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 2152 GCGTATACCAACAACACTGGGTAGTCAGATATACAAGGTAAAGGACCTGGATAATCGA 2211
 701 GlyLeuSerArgThr 705
 2212 GGCTTGTCAAGACA 2226

RESULT 15
 CQ492624

LOCUS CQ492624 2760 bp DNA linear PAT 30-JAN-2004
 DEFINITION Sequence 24491 from Patent WO0160860.
 ACCESSION CQ492624
 VERSION CQ492624.1 GI:41458243
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Schlegel, R., Endege, W.O. and Monahan, J.E.
 TITLE Genes differentially expressed in human prostate cancer and their use
 JOURNAL Patent: WO 0160860-A 24491 23-AUG-2001;
 FEATURES Millennium Predictive Medicine, Inc. (US)
 source
 1. 2760
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,628-305 Length: 2760
 Score: 3631.00 Matches: 698
 Percent Similarity: 98.87% Conservative: 1
 Best Local Similarity: 98.73% Mismatches: 6
 Query Match: 98.21% Indels: 2
 DB: 6 Gaps: 0
 US-10-028-384-12 (1-705) x CQ492624 (1-2760)
 QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 105 ATGACTAAGTTGGATTGTTGGATTGCTTATGAGAGAGCAGGACACACTTTTGAAGCTT 164
 QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu-PheAlaValLeuAr 40
 DB 165 CTCATTCTCTCAATGGCTGCTGTTATTCCTTCTCCATCGTCTGTTTGTCTGCTGAG 224
 QY 40 gPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLe 60
 DB 225 ATTGAAGTGTATCCATGAGTTTGAATGCGTACCTTTAATATCGACTACACAGTTCT 284
 QY 60 uAlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTyrTyrProLe 80
 DB 285 GGCTGAGGAGGGGTTTATAAATTCATAACCTGGTTTGAATGACCGAGCTGGTACCCCTT 344
 QY 80 uGlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTy 100
 DB 345 GGACAGCAATCTGGAGGACAAATTTACCCAGGTTTAAATGATCACCTCTGCTGCATCTA 404
 QY 100 rHisValLeuHis-PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuA 120
 DB 405 CCATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGCTGTGTGTCTCTGG 464
 QY 120 lAproLeuPheSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysA 140
 DB 465 CCCCTCTCTCTCTCTCCACCAUUGTCACTGATCCACCTTACCAAGAGCTCAAGG 524
 QY 140 sAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
 DB 525 ATCAGGGGCTGGGCTCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 584
 QY 160 rSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 180
 DB 585 GATCTGTGGCTGCTCCTATGATATGAAGAGGATTGCGCATCTTTTGCATCTACTACCT 644
 QY 180 YrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaLysCysAlaL 200
 DB 645 ACTACATGTGATCAAGGAGCAGTAAGACTGGTTCCATCTGTTGGGAGGAGTAAAGTGTGCC 704
 QY 200 euAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIleP 220

Db 705 TTGCTTATTCTACATGCTCTCGTCAATGGAGGTTATGTTCCTGATCACTTAATTC 764
Qy 220 rleuHlvalLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlar 240
Db 765 CTCTCCACGTCCTCGTCTGATGCTACAGCCGCTTTCTCTCACCGGATCTATGTGGCCT 824
Qy 240 YrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheG 260
Db 825 ACTGTACTGTTACTGCTGGGCACTATATCTTCTATGCAAGTCTCTTTTGGGTTTC 884
Qy 260 InProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleH 280
Db 885 AGCCTGTCTCTTTCATACAGACACATGGCAGCCTTTGGGGTCTTTGGTCTCTGCGACATCC 944
Qy 280 isAlaPheValAspTyrLeuArgSerIleValLeuPheValLeuValLeuPheA 300
Db 945 ATGCTTTTGGATTACCTCGGACGCAAGTTGAATCCACAACTTTGAAGTTCTTTTTC 1004
Qy 300 rGserValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuT 320
Db 1005 GGAGCGTCATCTCTCTGTTAGGCTTTGTCTCTCACCGTGGGAGCTCTCCATGCTGA 1064
Qy 320 hrGlyIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaL 340
Db 1065 CAGAAATATCTCCCTGGACGGGGGCTTCTACTCGTGTGGATCCCTCTTATGCTA 1124
Qy 340 ysAsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerT 360
Db 1125 AGAACACATCCCATCATTTCTCTGTCTGATGATCAGCATCAGCCACACACTGGTCTCAT 1184
Qy 360 YrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheS 380
Db 1185 ACTATTGTGACCTGACGCTCTCTGCTTCATGTTTCCAGTTGGCCTCTATTACTGCTTA 1244
Qy 380 erAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheS 400
Db 1245 GCAACCTGCTGATGCCGATTTTATCATCATGATGTTGGTGACACCATGATCTTTT 1304
Qy 400 erAlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyI 420
Db 1305 CAGCTGTAAATGGTGGCTCTAATGCTAGTGTGGACCTGTATGTCATTCCTCTGGCA 1364
Qy 420 leGlyValSerGlnValLeuSerThrTyrMetIleAsnLeuAspIleSerArgProAspL 440
Db 1365 TTGGAGTCTCCAGGTGCTGCTCCACATACATGAAGAATCTGGACATAAGTCTGCCAGACA 1424
Qy 440 ysLysSerLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyW 460
Db 1425 AGAAGAGCAAGCAAGCAAGATTCCACCTTACCTTATTAAGAATGAAGTGCAAGTGGA 1484
Qy 460 etIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrS 480
Db 1485 TGATACGTGTCATGGCTTTCTTCTCATCACCTTACACCTTTTCATCAACCTGGGTGACCA 1544
Qy 480 erGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgI 500
Db 1545 GTGAGGCTTACTCTTCTCGCTCCATTTGTTCTGCTGCTCCCGTGGTGGGATGTCAGTAGGA 1604
Qy 500 leIlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspA 520
Db 1605 TCATATTGATGACTTCCGAGAGACATATTATTGGCTTCGTTCATAATACTCCAGAGGATG 1664
Qy 520 lalysValMetSerTrpTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrI 540
Db 1665 CGNAGGTCTATGCTGGGATTAATGGCTATCAGATTACAGCTATGGCAACCGCAACAA 1724
Qy 540 leLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetA 560
Db 1725 TTTTAGTGACATAACACATGAATAATACCATATTCTCGAGTAGGGCAGGCAATGG 1784
Qy 560 laSerThrGluGlyAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuV 580

Db 1785 CGTCCACAGAGGAAAAAGCCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGTGCTGG 1844
Qy 580 alIlePheGlyGlyLeuThrGlyTyrSerSerAspPheAsnLysPheLeuTrpMetV 600
Db 1845 TCATTTTGGAGGCTCCTGCTGATGATGATCAACAAGTTTCTTTGGATGG 1904
Qy 600 alArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrP 620
Db 1905 TCCGATTGGAGGGAGCAGATACAGCAACATATCAAGGAGATGACTATTATATCTC 1964
Qy 620 toThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetT 640
Db 1965 CAACCTGGGGAGTTCGCTGTGACCGTGAAGTTCCTCCAGTCTCTCAACTGCCTCATGT 2024
Qy 640 YrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaIleArgProProGlyP 660
Db 2025 ACAAGATGTGTACTATCTGCTTTGGACAGGTTTACACAGAAAGCCAGCGCTCTCCAGGCT 2084
Qy 660 heAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluG 680
Db 2085 TTGACCGTGTCCGAATGCTGAGATTGGGAATAAGACTTTGAGCTTGATGCTCTGGAGG 2144
Qy 680 luGlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnA 700
Db 2145 AAGCATATACCACAGAACATTTGGCTGGTCAAGATATACAGGTAAGGACCTGGATATC 2204
Qy 700 rGlyLeuSerArgThr 705
Db 2205 GAGGCTTGTCAAGGACA 2221

Search completed: December 15, 2004, 06:01:36

Job time : 7905.52 secs

Blank
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:28:05 ; Search time 812.511 Seconds
(without alignments)
4554.822 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 3697

Sequence: 1 MTKGFLRLSYEKQDTLLKL.....HMLVRIYKVDLNRGLSRT 705

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10028384/runat 14122004 131518 8007/app query.fasta_1.3740
-DB=N_Geneseq_23Sep04 -QMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MAFIX=blosun62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 -CGN 1 1 1986 -runat 14122004 131518 8007 -NCPU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N_Geneseq_23Sep04: *
1: Geneseqn1980s: *
2: Geneseqn1990s: *
3: Geneseqn2000s: *
4: Geneseqn2001as: *
5: Geneseqn2001bs: *
6: Geneseqn2002as: *
7: Geneseqn2002bs: *
8: Geneseqn2003as: *
9: Geneseqn2003bs: *
10: Geneseqn2003cs: *
11: Geneseqn2003ds: *
12: Geneseqn2004s: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3697	100.0	2472	10 ADD94793	Add94793 Human ITM
2	3697	100.0	2472	10 ADH28844	Adh28844 Human chr
3	3633	98.3	3094	10 ADD94791	Add94791 Mouse ITM
4	3631	98.2	2760	5 ABV24502	Abv24502 Human pro
5	2903.5	78.5	2855	4 ABL02795	Ab102795 Drosophila
6	2602.5	70.4	6153	4 ABL02794	Ab102794 Drosophila

7	2263.5	61.2	1371	12 ADP28508	Adp28508 Human sec
8	2247	60.8	2481	10 ADD94783	Add94783 Human SIM
9	2237	60.5	2710	10 ADD94785	Add94785 Mouse SIM
10	2177.5	58.9	2417	10 ADD94789	Add94789 Drosophila
11	2177.5	58.9	2699	4 ABL13247	Ab113247 Drosophila
12	2103	56.9	4922	4 ABL13246	Ab113246 Drosophila
13	2070.5	56.0	2256	6 ABZ32036	Abz32036 candida a
14	2010	54.4	2232	8 ABT20789	Abt20789 Aspergill
15	1930	52.2	2466	10 ADB69300	Adb69300 C. neofor
16	1920	51.9	2733	10 ADD94787	Add94787 Yeast STT
17	1867.5	50.5	2000	10 ACC61293	Acc61293 Gene sequ
18	1867.5	50.5	2000	10 ADK63659	Adk63659 Disease t
19	1851	50.1	2603	8 ABT20191	Abt20191 Aspergill
20	1851	50.1	3969	8 ABT17781	Abt17781 Aspergill
21	1851	50.1	4603	8 ABT19595	Abt19595 Aspergill
22	1722	46.6	1848	8 ABT18969	Abt18969 Aspergill
23	1682	45.5	2882	10 ADB69539	Adb69539 C. neofor
24	1682	45.5	4738	10 ADB69178	Adb69178 C. neofor
25	1680	45.4	1969	8 ABT18375	Abt18375 Aspergill
26	1273	34.4	1664	4 AAK94164	Aak94164 Human ful
27	1273	34.4	1664	12 ADL30661	Adl30661 Full leng
28	1241.5	33.6	1209	4 AAB33284	Aab33284 Human col
29	1241.5	33.6	1209	6 ABL89850	Ab189850 Human pol
30	1241.5	33.6	1543	2 AAX85055	Aax85055 Human sec
31	1241.5	33.6	1543	8 ACD18981	Act18981 Novel hum
32	1241.5	33.6	1543	12 ADG78372	Adg78372 Human sec
33	1241.5	33.6	1543	12 ADN60563	Adn60563 Human sec
34	890.5	24.1	2510	5 AAF93772	Aaf93772 Human CDN
35	890.5	24.1	2546	2 AAV44866	Aav44866 Clone CT5
36	890.5	24.1	2546	5 AAF98433	Aaf98433 Human CDN
37	880.5	23.8	2537	4 AAD58289	Aad58289 Human sec
38	878.5	23.8	2284	4 AAH18021	Aah18021 Human CDN
39	875.5	23.7	2547	4 AAD08315	Aad08315 Human sec
40	875.5	23.7	2660	6 ABQ54750	Abq54750 Human ova
41	870.5	22.2	1114	4 AAH99794	Aah99794 Human pro
42	805	21.8	487	3 AAL16208	Aal16208 Human col
43	802	21.7	455	6 ABL37867	Ab137867 Human col
44	793.5	21.5	787	4 AAH07526	Aah07526 Human CDN
45	792	21.4	440	6 ABK27622	Abk27622 Human col

ALIGNMENTS

RESULT 1

ADD94793
ID ADD94793 standard; DNA; 2472 BP.

XX AC ADD94793;

XX DT 29-JAN-2004 (first entry)

XX DE Human ITM1 gene sequence.

XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;
major histocompatibility complex; human leukocyte antigen; HLA;
cytotoxic; immunosuppressive; antisense therapy; gene therapy; cancer;
lung cancer; intestine cancer; sarcoma; prostate cancer;
testicular cancer; breast cancer; melanomas; pancreatic cancer;
haematological cancer; immune response; lymphoid cell proliferation;
autoimmune disease; transplant rejection; SIMP-derived peptide; human;
gene; ds; ITM1.

XX OS Homo sapiens.

XX PN WO2003054008-A2.

XX XX 03-JUL-2003.

XX PF 18-DEC-2002; 2002WO-CR001967.

XX PR 20-DEC-2001; 2001US-00028384.

XX PA (COMP-) COMPATIGENE INC.

Db 1607 ATATTGATGACCTCCGAGAGCATATATTGGCTTCGTATATCTCCAGAGGATGG 1666
 Qy 521 LysValMetSerTrpTrpAspTyrGlyTyrGluIleThrAlaMetAlaAsnArgThrIle 540
 Db 1667 AAGGTCATGTCCTGGTGGGATTATGGCTATACAGATTACAGTATGGCAACCGACAAAT 1726
 Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1727 TTAGTGGACATAACACATGATGAATAATACCATATTTCTGAGTAGGCGAGCAATGGG 1786
 Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1787 TCCACAGAGGAAAAGGCTATGAGATCATGAGGAGCTGATGTCAGTATGCTGGTGC 1846
 Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 Db 1847 ATTTTGGAGGCTCCACTGGGTATCTCTGATGATATCAACAAGTTTCTTTGGATGGTC 1906
 Qy 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrThrPro 620
 Db 1907 CGGATTGGAGGAGCAGATACAGGCAACATATCAAGGAGAAATGACTATTATCTCCA 1966
 Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1967 ACTGGGAGTTCCTGGTGGACCGTGAAGTTCTCCAGTGTCTCAACTGCCTCATGTAC 2026
 Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 2027 AAGATGTGTACTATCTGCTTTGGACAGGTTTACACAGAACCCAGCGTCTCCAGGCTTT 2086
 Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 2087 GACCGTGTCCGAATGCTGAGATTGGCAATAAAGACTTTGAGCTTGATGTCCTGGAGAA 2146
 Qy 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2147 GGCTATACCAAGAACATGCTGCTGCTGAGGATATACAGGTAAGGACCTGGATATCGA 2206
 Qy 701 GlyLeuSerArgThr 705
 Db 2207 GCTTGTGTCAGGACA 2221
 RESULT 2
 ADH28844
 ID ADH28844 standard; DNA; 2472 BP.
 XX
 AC ADH28844;
 XX
 DT 11-MAR-2004 (first entry)
 DE Human chronic myelogenous leukaemia (CML) gene marker #112.
 XX
 KW ds; chronic phase chronic myelogenous leukaemia; CP-CML;
 KW blast crisis CML; BC-CML; human; chronic myelogenous leukaemia;
 KW gene marker.
 XX
 OS Homo sapiens.
 XX
 FN US2003104426-A1.
 XX
 PD 05-JUN-2003.
 XX
 PF 14-JUN-2002; 2002US-00171581.
 XX
 PR 18-JUN-2001; 2001US-0298914P.
 XX
 PA (LINS//) LINSLEY P S.
 PA (MAOM//) MAO M.
 PA (DAIH//) DAI H.
 PA (HEY//) HE Y.
 PA (RADI//) RADICH J P.
 XX
 PI Linsley PS, Mao M, Dai H, He Y, Radich JP;

XX WPI; 2003-787045/74.
 DR
 XX
 PT Classifying cell sample as chronic phase chronic myelogenous leukemia or
 PT blast crisis chronic myelogenous leukemia by detecting difference in
 PT expression of genes corresponding to the markers such as X15415, U99436.
 XX
 PS Disclosure; SEQ ID NO 112; 31pp; English.
 XX
 CC The invention relates to a method of classifying a cell sample as chronic
 CC phase chronic myelogenous leukaemia (CP-CML) or blast crisis CML (BC-
 CC CML). The method is useful for classifying a sample as CP-CML or BC-CML.
 CC The present sequence represents a human chronic myelogenous leukaemia
 CC (CML) gene marker used to distinguish blast crisis CML from chronic phase
 CC CML.
 XX
 SQ Sequence 2472 BP; 566 A; 568 C; 583 G; 755 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 2472
 Score: 3637.00 Matches: 705
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0
 US-10-028-384-12 (1-705) x ADH28844 (1-2472)
 Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 Db 107 ATGACTAAGTTTGGATTTTTCGATTTCTCTATGAGAGCAGGACACACTTTTGAAGCTT 166
 Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 Db 167 CTCTATTCGTCAATGGCTGCTGTATTTCTCTCCACTCGTCTGTTGCTGCTCTGAGA 226
 Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 Db 227 TTTGAAAGTGTATTCATGATGATTTGATCCGTACTTTAATTATCGGACTACCAAGTTCCTG 286
 Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
 Db 287 GCTGAGGAGGGGTTTATAAATTCATTAATCTGGTTGATGACCGAGCCTGGTACCTTTG 346
 Qy 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 347 GGACGAATCATTCGAGGAGCAATTTACCCAGGTTTAAATGATCACTCTCTGCAATCTAC 406
 Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db 407 CATGTACTCCATTTTTCACATCACATCGACATTCGGAAATGCTGTGTGTCTGTCGCC 466
 Qy 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 Db 467 CCT 526
 Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
 Db 527 GCAGGGGCTGGGCTTCCTGCTGCTGCATGATTCGTAGTTCCTGGAATATCTCCCGA 586
 Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 Db 587 TCTGTGGCTGGCTCTCTATGATATGAAGGGAATGCCATCTTTTGGATGCTACTCACTAC 646
 Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
 Db 647 TACATGTGGATCAAGGAGTAAAGACTGGTTCATCTGTGGGAGCTAAGTGTGCTCTT 706
 Qy 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIlePro 220
 Db 707 GCTTATTCTTACATGGTCTCGTCATGGGAGGTTATGTCTCTGATCACTCACTAACTTCT 766
 Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240

Db 767 CTCCACGCTCTCGTCTGATGCTCACAGGCGGTTTCTCTCACCGGATCTATGTGCGCTAC 826
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db 827 TGTACTGTTACTGCTGGTGGTACTATATCTTTAGGAGATCTCTTTGTGGGTTTCAG 886
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 Db 887 CCTCTCTCTTTCATCAGACACATGGCAGGCTTTGGGGTCTTTGGTCTCTGCCAGATCCAT 946
 QY 281 AlaPheValAspTyrIleuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db 947 GCCTTTGGGATACCTGGCAGCAAGTTGAATCCACACAAATTTGAAGTCTTTTCGG 1006
 QY 301 SerValIleSerLeuValGlyPheValLeuThrValGlyAlaLeuLeuMetLeuThr 320
 Db 1007 AGCGTCATCTCTGTGTAGGCTTTGTCTCTCACCGTGGAGCTCTCTCATGTCTGACA 1066
 QY 321 GlyLysIleSerProTptThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 Db 1067 GGAATAATATCTCCCTGGACGGGGGCTTCTACTCACTGCTGGATCCCTCTTATGCTAAG 1126
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 Db 1127 AACAAACATCCCATCATTTGCTGTCTGAGCATCAGCCACCAACCTGTCTCTCATAC 1186
 QY 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 Db 1187 TATTTTGACCTGAGCTCTCTGCTTCATGTTTCCAGTTGGCTCTTATTAAGCTTTAGC 1246
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 Db 1247 AACCTGTCTGATGCCGGATTTTATCATCATGATGTTGTTGACCAAGCATGTACTTTCA 1306
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 Db 1307 GCTGTATGTTGGTCTTATGCTATGTTGGACCTGTTATGAGCATTTCTCTTGGCAT 1366
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 Db 1367 GGAGTCTCCAGAGTGTCTGCATACATCAATCAATGATGATGATGATGATGATGATGAT 1426
 QY 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 Db 1427 AAGAGCAAGAGCAGCAGGATTCACCTACCTATTAAGATTGAATGGCAGTGGATG 1486
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThrSer 480
 Db 1487 ATACTGTCTATGCTTTCTTCTCATCACCTACCTACCTTTTCAATCAACCTGGTACCA 1546
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyLysAspGlySerArgile 500
 Db 1547 GAGGCTTACTTCTCTCGTCCATTTGATCTATCTGCCCGTGGTGGGATGGCAGTAGATC 1606
 QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 Db 1607 ATATTTGATGACTCCGAGAGCATATATTGGCTTCTGTCATAATCTCCAGAGGATGCG 1666
 QY 521 LysValMetSerTrpTptAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 Db 1667 AAGTCTATGCTCTGTTGGGATTTAGCTATCAGATTACAGCTATGGCAACCGACAAAT 1726
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1727 TTAGTGACATAAACACATGAAATAATACCATATTCTCGAGTAGGCGAGCAATGGCG 1786
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1787 TCCACAGAGAAAGCCTATGAGATCATGAGGAGCTCGATGTGAGCTATGTGTGCTGTC 1846
 QY 581 IlePheGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600

Db 1847 ATTTTGGAGGCTCCTCAGTGGTATTCTCTGATGATATCAACAGTTTCTTTGGATGGTC 1906
 QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1907 CGGATTTGGAGGAGGACACAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCCA 1966
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1967 ACTGGGAGTTTCCGTGTGGACCGTGAAGTTCTCCAGTGTGCTCAACTGCCCTCATGTAC 2026
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 2027 AAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAGCCAAAGCGTCTCTCAGGCTTT 2086
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 2087 GACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGCTCTGGAGGAA 2146
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2147 GCCTATACACAGAACATTGGCTGTGTCAGGATATACAAAGGTAAAGGACCTGGATATCGA 2206
 QY 701 GlyLeuSerArgThr 705
 Db 2207 GGCTTTGTCAAGGACA 2221
 RESULT 3
 ID ADD94791 standard; DNA; 3094 BP.
 XX ADD94791;
 AC ADD94791;
 DT 29-JAN-2004 (first entry)
 XX Mouse ITM1 gene sequence.
 DE source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytotoxic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 KW murine; gene; ds; ITM1.
 XX Mus musculus.
 OS WO2003054008-A2.
 XX 03-JUL-2003.
 XX 18-DEC-2002; 2002WO-CA001967.
 XX 20-DEC-2001; 2001US-00028384.
 XX (COMP-) COMPATIGENE INC.
 XX Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 XX P-PSDB; ADD94792.
 XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 XX acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 XX or breast cancer, or for suppressing an immune response in an autoimmune
 XX disease.
 XX Disclosure; SEQ ID NO 9; 66pp; English.
 PS This invention relates to a novel isolated or purified human protein.
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein

1

QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspThrThrPro 620
 Db 1912 CGGATTGGAGGAGACAGAGAGAGAGAGACACATTAAGGAGAGTACTACTACTCT 1971
 QY 621 ThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1972 ACTGGGAATTCGGTGTGATGAGGGTCTCCCGTCTGCTCACTGCCCTTAAGTAC 2031
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
 Db 2032 AAAATGTGTACTACCGCTTTGGCGAGGTCTACACAGAGCAAGCGTCCACAGGCTT 2091
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 2092 GACCGTGTTCGAATGCTGAGATTGGTAATAAGACTTTGAGCTGTGCTGGAGGAA 2151
 QY 681 GlyTyrThrGluHisTyrLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2152 GCGTATACACAGAACACTGGCTAGTCAGGATATACAGGTAAGGACCTGGATAATCGA 2211
 QY 701 GlyLeuSerArgThr 705
 Db 2212 GCGTTGTCAAGGACA 2226
 RESULT 4
 ID ABV24502 standard; cDNA; 2760 BP.
 XX AC ABV24502;
 XX XX
 DT 16-SEP-2002 (first entry)
 XX XX
 DE Human prostate expression marker cDNA 24493.
 XX XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX Homo sapiens.
 OS
 XX WO2001:06860-A2.
 PD
 PD 23-AUG-2001.
 PF
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 XX Schlegel R, Endege WO, Monahan JE;
 PI WPI; 2001-662795/76.
 XX
 DR Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 PS Claim 1; Page 4632-4633; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 2760 BP; 688 A; 645 C; 658 G; 769 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 2760
 Score: 3631.00 Matches: 698
 Percent Similarity: 98.87% Conservative: 1
 Best Local Similarity: 98.73% Mismatches: 6
 Query Match: 98.21% Indels: 2
 DB: 5 Gaps: 0
 US-10-028-384-12 (1-705) x ABV24502 (1-2760)
 QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
 Db 105 ATGACTTAAGTTTGGATTTTTGGATTGCTCTATGAGAAGCAGGACACACTTTTGAAGCTT 164
 QY 21 LeuLeuLeuSerMetAlaValLeuSerPheSerThrArgLeu-PheAlaValLeuAr 40
 Db 165 CTCATTCTGTCATGCTGCTGATATATCCCTCTCCACTCGCTCTGTTTGTCTGCTCTGAG 224
 QY 40 gPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLe 60
 Db 225 ATTGGAAGTGTATCCATGAGTTTGTGCTGCTCTTAAATATATCGGACTACCAAGTTTCT 284
 QY 60 uAlaGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTrpProLe 80
 Db 285 GCGTGGAGGGGGTTTATAAATCCATACTGTTTGTGATGACCGACCTGGTACCTTTT 344
 QY 80 uGlyArgIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 345 GGGAGCAATCATTTGAGGAGCAAAATTTACCAAGTTTAAATGATCACTCTGCTGCAATCTA 404
 QY 100 rHisValLeuHis-PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuA 120
 Db 405 CCATGTAATCTCCATTTTTCACATCCACATCGACATTCGGAATGCTGTGTCTCTCGG 464
 QY 120 laProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuLeuLysGluLeuLysA 140
 Db 465 CCCCTCTCTCTCTCTCCATCCACCATCGTCAGCTACCACTTACCAAAAGAGCTCAAGG 524
 QY 140 spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
 Db 525 ATGCAGGGGCTGGGCTCTTGTCTGCTGCCATGATGCTGTAGTCTCTGGATATATCTCCC 584
 QY 160 rGSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrT 180
 Db 585 GATCTGGCTGGCTCTCTATGATATATGAAGGATGTCATCTTTTGCATGCTACTACCT 644
 QY 180 yTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaL 200
 Db 645 ACTACATGTTGATCAAGGAGGAGTAAAGACTGTTTCCATCTGTTGGGAGCTAAGTGTGCC 704
 QY 200 euAlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuLeuLeuLeuLeuP 220
 Db 705 TTGCTTATTTTACATGCTGCTCATGGGAGGTTATGTTCTCTGATCAACTTAATTC 764
 QY 220 roleuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaL 240
 Db 765 CTTCTCCAGTCTCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 824
 QY 240 yTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaL 260
 Db 825 ACTGTAATGTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
 QY 260 laProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuLysGluLeuH 280
 Db 885 AGCTGTCTCTTTCATCAGAGACATGGGAGCTTTGGGGTCTTTGGTCTCTGCGAGATCC 944

QY 280 isAlaPheValAspTyrLeuArgSerTyrLeuAsnProGlnInPheGluValLeuPheA 300
DB 945 ATGCGTTTGTGGATTACCTGGCGAGCAAGTTGAATCCACAACTTTGAAGTTCTTTTC 1004
QY 300 rGSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuT 320
DB 1005 GGAGGGGTATCTCTCTGTAGAGCTTTGCTTCTCACCGTGGAGCTCTCTCATGCTGA 1064
QY 320 hrGlyIleSerProThrTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaL 340
DB 1065 CAGGAAAAATATCTCCCTGGACGGGGCTTTCTACTCGCTCGTGGATCCCTCTTATGCTA 1124
QY 340 ysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerT 360
DB 1125 AGAACACATCCCCATCATTTGCTTCTGTCTGAGCATCAGCCACAACTGGTCTCTCAT 1184
QY 360 yTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheS 380
DB 1185 ACTATTTTACCTGAGCT 1244
QY 380 erAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheS 400
DB 1245 GCAACCTGTCTGATGCCGGATTTTATCATCATGTATGTGTGACCGCATGTATCTTT 1304
QY 400 erAlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyI 420
DB 1305 CAGCTGTAAATGCTGCTTAATGCTAGTGTGGACCTGTATGTGCAATCTCTCTGGCA 1364
QY 420 leGlyValSerGlnValLeuSerThrTyrMetTyrAsnLeuAspIleSerArgProAspL 440
DB 1365 TTGGAGTCTCCAGGTGTCTCCACATACATGATGATGATGATGATGATGATGATGATG 1424
QY 440 ysLysSerIleGlyGlnInAspSerThrTyrProIleLysIleGluValAlaSerGlyM 460
DB 1425 AGAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1484
QY 460 erIleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValTrS 480
DB 1485 TGATACTGTGTATGGCTTTCTTCTCATCATCATCATCATCATCATCATCATCATCAT 1544
QY 480 erGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgI 500
DB 1545 GTGAGGCTACT 1604
QY 500 leilePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspA 520
DB 1605 TCATATTTGATGACATTCGAGAGCATATATATGCTTGTCTATATATATATATATATAT 1664
QY 520 laLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrI 540
DB 1665 CGAAGGTATGCTCTGTGGATTTATGGCTATCAGATTACAGCTATGGCAACCGAACAA 1724
QY 540 leLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetA 560
DB 1725 TTTTGTAGTGACATPACACATGATGATATATACCATATTTCTCGATAGGAGCAGCATGG 1784
QY 560 laSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuV 580
DB 1785 CGTCACACAGAGAAAGCTATGAGATCATGAGGAGCTGATGTCTGATGTCTGCTGCTG 1844
QY 580 aIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetV 600
DB 1845 TCATTTTGGAGGCTCTCTGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1904
QY 600 alArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrP 620
DB 1905 TCCGATTTGGAGGAGCAGATACAGCAACATATCAAGGAGATGATGATATATATCTC 1964
QY 620 roThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetT 640
DB 1965 CAACTGGGAGTTCCTGCTGACCGCTGAGGTTCTCCAGTGTCTCTCACTGCTCTATGT 2024
QY 640 yzLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyP 660

DB 2025 ACAAGATGTGTACTATCGTTTGGACAGGTTTACACAGAACCAAGCGTCTCCAGGCT 2084
QY 660 heAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluG 680
DB 2085 TTGACCGTGTCCGAAATGCTGAGATTGGGATTAAGACTTTGAGCTTGTATGCTTGGAGG 2144
QY 680 luGlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnA 700
DB 2145 AAGCATATACACAGAACATTGGCTGTGTCAGGATATACAGGTAAAGGACCTGGATAATC 2204
QY 700 rGlyLeuSerArgThr 705
DB 2205 GAGCTTGTCAAGACA 2221
RESULT 5
ID ABL02795 standard; cDNA; 2855 BP.
XX ABL02795;
AC ABL02795;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 2867.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX MPI; 2001-656860/75.
XX P-PSDB; ABB58692.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 2867; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 2855 BP; 661 A; 750 C; 724 G; 720 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.21e-284 Length: 2855
Score: 2903.50 Matches: 536
Percent Similarity: 87.02% Conservative: 74
Best Local Similarity: 76.46% Mismatches: 84
Query Match: 78.54% Indels: 7
DB: 4 Gaps: 1

RESULT 6

ABL02794
 XX ABL02794 standard, cDNA; 6153 BP.
 AC ABL02794;
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2864.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-VAR-2001; 2001WO-US009231.
 PF
 XX 23-MAR-2000; 2000US-0191637P.
 PR
 XX 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEXE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EM;
 PI
 XX WPI: 2001-656860/75.
 DR
 XX P-PSDB; ABB58691.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 2864; 21pp + Sequence Listing; English.
 CC
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU130511), expressed DNA
 CC sequences (ABU1840-ABU16175) and the encoded proteins (ABB57737
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 6153 BP; 1692 A; 1334 C; 1299 G; 1828 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,62e-253 Length: 6153
 Score: 2602.50 Matches: 524
 Percent Similarity: 60.24% Conservative: 70
 Best Local Similarity: 53.14% Mismatches: 79
 Query Match: 70.39% Indels: 315
 DB: 4 Gaps: 5
 US-10-028-384-12 (1-705) x ABL02794 (1-6153)
 QY 30 SerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValIleHisGluPheAsp 49
 DB 1890 GCATTGGCACACCGCTTCTCTGTCTGCGATTCGAAAGCGTAATCCATGAGTTTCGAT 1949
 QY 50 ProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGlyPheTyrLysPheHis 69
 DB 1950 CCGTACTTCAACTACCGCACACCGGGTTCTGCGGAGCAGGGCTTTTACAGTTCCAC 2009
 QY 70 AsnTyrPheAspArgAlaTyrTyrProLeuGlyArgIleGlyGlyThrIleTyr 89
 DB 2010 AACTGGTTTCGATGACCGCGCTGGTATCCCTGGCGCGCATCATCGCGGCACATCAT 2069
 QY 90 ProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHisPhePheHisIleThr 109

2070 CCGGGCTGATGCTCACTCGGCGGCTGTACCGCTGTGCTGCTCAATGTGACC 2129
 QY 110 IleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerSerPheThrSerIle 129
 DB 2130 ATGCACATACGGAACGTCGCTCTCTCTGCGCCCTTCTTCTCTCGCTGACACGCTG 2189
 QY 130 ValThrTyrLeuLeuThrLysGluLeu- 138
 DB 2190 GTGACCTACGCCCTCACAAAGGAGATACACGTAAAGCCACCGTCCACCGCTTCTTACTTT 2249
 QY 138 ----- 138
 DB 2250 TTTTATTATCTCTTCGAGGTTAAATTCGACGGTCTTATGATCCCTTTATCC 2309
 QY 139 -----LysAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGly 156
 DB 2310 CTTTTTAGAGC-ACTGGAGCTGGACTGGTGGCGCGCTTGTATATCCATCGTTCCCGG 2368
 QY 157 TyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMet 176
 DB 2369 TATATCTCGATCCGTGGCGGATCGTAGCAATGAAGGCATCGCCATTTCTGCATG 2428
 QY 177 LeuLeuThrTyrTyrMetTyrIleLysAlaValLysThrGlySerIleCysTrpAlaAla 196
 DB 2429 CTCCTCACCTACTATTTGTGATCAAGCGGTAAAGACGGGCACGATCTTTTGGTCGGCT 2488
 QY 197 LysCysAlaLeuAlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIle 216
 DB 2489 ATGTCGCATTTGGCTACTTCTATATGCTCTCTCTGGGGTGGCTATGCTTCTTGAT 2548
 QY 217 AsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIle 236
 DB 2549 AACCTAATCCCGCTGCACGCTGGCGGTGATCATCACGGACGTTTCTCGCACAGATC 2608
 QY 237 TyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPhe 256
 DB 2609 TACATAGCTACACACGCTATCTACTGCTGGCACCATTCTGTCGATGAGATCTCGTT 2668
 QY 257 ValGlyPheGlnProValLeuSerSerGluHisMetAlaGly- 270
 DB 2669 GTGGGATTCACCCATCCAGAGCTCGAACACATGCT-GGTGAGTGTAACTACTTTGTG 2727
 QY 270 ----- 270
 DB 2728 TAATGCTGTGAGTGTAGCGATGTGTATCTGTTGTGATGATACATTTTAAATTTTGA 2787
 QY 270 ----- 270
 DB 2788 TAGGACGAGTATGAACATATGATGATGACCAAAACCTAGTGAATTGGTCACTAAAGAA 2847
 QY 270 ----- 270
 DB 2848 CTCTAATAAAGAATTAATCTGCTATATCGAAAGAGGAGTGAATAATCGTATTTTCTC 2907
 QY 271 -----PheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeuArg 287
 DB 2908 TTGCAGGCACTGGGCAACCTTTGGCTGTGCCAGATTTCACGCTTTCGTCGACTATCTGCGC 2967
 QY 288 SerLysLeuAsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuValGly 307
 DB 2968 TCGCGCATTCACAGGATCACTTCGATCTGCTCTTCAAGACGTTGGTTTCCAGTGTGTTG 3027
 QY 308 PheValLeuLeuThrValGlyAlaLeuLeuMetLeuThrGlyLysIleSerProTyrThr 327
 DB 3028 ACTGTGTGTTCGTCGGGTACCTTCCTCACGCTTACCGGGAAGTCTCCCTCGGACC 3087
 QY 328 GlyArgPheTyrSerLeuLeuAspProSerTyrAlaLysAsnIleSerIleAla 347
 DB 3088 GGCAGATTTTACTCGCTACTAGATCCATCTATGCCAAGATCCATCCATCCATCATCGCC 3147
 QY 348 SerValSerGluHisGlnProThrThrTyrSerTyrTyrPheAsp- 363

Db 3148 TCCGATATCGGAGCATCAGCCTACATCCTCGTGGTCTTTCTATTGATCTGCAGTAGGT 3207
 QY 363 ----- 363
 Db 3208 AAACCACTATCTCGTAGGAATCTGTAGATGCTCTCAACTTAACAGTGTGCTTTATC 3267
 QY 363 ----- 363
 Db 3268 ATATCTATTTTGTGTTAAATAATCTTACTACAAACAAATTCGGTTTATTTTAAACCTAT 3327
 QY 363 ----- 363
 Db 3328 CCTAATTACACATGTTCTCGGAGGATGAGAACTTAACACAGATGGAATGATATGCGCC 3387
 QY 363 ----- 363
 Db 3388 TTAATGCAATTGAGTATTACCTTGCATCTTTGATCGCTGAGCGAGTTAAATAACATGACC 3447
 QY 364 ----- 364
 Db 3448 CCTTTTCACATAACAATCAATATATAGTTAATTCATTCGCTTTTTCACAAATTCCTTT 3507
 QY 364 uGlnLeuValPheMetPheProValGlyLeuTyrCysPheSerAsnLeuSerAs 384
 Db 3508 GCAGATCCTGGTCTCTCTCCGCTGCGCTTTATTTCTGCTTTTCCAAAGCTGAGCGA 3567
 QY 384 pAlaArgilePheileileMetTyrGlyValThrSerMetTyrPheSerAlaValMetVa 404
 Db 3568 CTCGAACATCTTATTTATTTGATGGCTCACCAGTATTATTTTCGCGGAGTATGATGT 3627
 QY 404 lArgLeuMetLeuValLeuAlaProValMetSerileLeuSerGlyileGlyValSerGl 424
 Db 3628 GCGTCTAATGCTGCTCTCGCGCGCGTGTGTGTCTATCCGGAATGCGCAATTTCGCA 3687
 QY 424 nValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLysLysSer---- 442
 Db 3688 TCTCTGGCCAAAGTATATCAAGAGCGTTGATCCGGAAGCTCATCGAGCGAGGAGTAGA 3747
 QY 443 -----LysLysGlnGlnAspSerThrTyrProIleLysIleGluValAl 457
 Db 3748 AAGCAAGCGCAGCACCAAGAGTGGACACAGCAGCAGCGGAGCGCTGAAGAGCGAGGTGCG 3807
 QY 457 aserGlyMetileLeuValMetAlaPhePheLeuileThrTyrPheHisSerThrTr 477
 Db 3808 CATTTGGATTCTGGGTCTCATCTCTGATGCTTATAGTGTACAGCTGCACCTGCACCTG 3867
 QY 477 pValThrSerGluAlaTyrSerSerProSerileValLeuSerAlaArgGlyGlyAspGl 497
 Db 3868 GGTCACTCGAGGCGCTTCTCTCGCCAGTATTGTTGAGTGCAGGTGCGCAGATGG 3927
 QY 497 ySerArgileilePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrPr 517
 Db 3928 CGGCCGATCATTTTCGATGACTTCGCGAGGCGCTACTACTGCTGCGATGACACTCC 3987
 QY 517 oGlu----- 518
 Db 3988 GGAGGTGAATGAAGTTTGTATTGCCATTTTGTGAGTGTCTCTCGCGAGTTAGCCATAG 4047
 QY 518 ----- 518
 Db 4048 AAATCAGAAACAGATTGTGTAATGAATAATTGTGGGCTTTTAAGATCGGAATTTTCA 4107
 QY 518 ----- 518
 Db 4108 TCTGGCTAAGTCGGCAACACAGAGTTTCGATCTTTAATGATGACTACTATTTGTAATTTCT 4167
 QY 518 ----- 518
 Db 4168 TGATAGGAGATGTAACATAATAGTACTATAAACACCAAGCTGACAAACAGATA 4227
 QY 518 ----- 518
 Db 4228 TCCCTTTTATTTGCAAACTTTTGTAAATGTCTGCACGTGACCGTTTCTGTAATGATTTT 4287

QY 519 ----AspAlaLysValMetSerTyrTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAs 537
 Db 4288 TCAGGAGGCTCGCAATATGTTCTGTGGGACTACGGCTACAGATAACGCCATGGCCAA 4347
 QY 537 nArgThrileLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGl 557
 Db 4348 TCGGACGATATTGGTGGATAACAATCTTGGAAACAACACATATATCGCGCGTTGGCCA 4407
 QY 557 nAlaMetAlaSerThrGluGluLysAlaTyrGluileMetArgGluLeuAspValSerTy 577
 Db 4408 GCGGATGGCTCTTCGGAGGAGAAAGCCACAGATAATGAGGAACTGGATGTGGACTA 4467
 QY 577 rValleuValillePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLe 597
 Db 4468 CDTTCTCGTGATTTTCGGAGGCGCTCAGTGGCTACTCATCGACGATATCAACAGTTTCT 4527
 QY 597 uTrpMetValArgileGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTy 617
 Db 4528 GTGGATGGTGGCAATTGGCGCAGCAGCATCGTGTGGCAGCATCCGGAAGGACTA 4587
 QY 617 rTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValleuLeuAsnCy 637
 Db 4588 CTATCGGCCAACGAGAGATTCCGAGTGGACAAGAGGGGCTCACCCACACTGTCTAATTG 4647
 QY 637 sLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgPr 657
 Db 4648 TTTGATGTACAGATGTCTCTACTATCGTTTGGCAAAATGTACACGGAAGTGGCAAGC 4707
 QY 657 oProGlyPheAspArgValArgAsnAlaGluileGlyAsnLysAspPheGluLeuAspVa 677
 Db 4708 CCAGGCGCTACGATCGAGTTCGTGCGCGGAGATCGGCAACAGGACTTTGAATGGATGT 4767
 QY 677 lLeuGluGluGlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLe 697
 Db 4768 CATGAGAGGGCGGTACACCCAGGAGCTGGCTGGTGGCATCTTACAAGGTTAAGGATCT 4827
 QY 697 uAspAsnArgGlyLeu 702
 Db 4828 GCCGAATCTGGAGTC 4843
 RESULT 7
 ID ADP28508 standard; DNA; 1371 BP.
 XX ADP28508;
 AC ADP28508;
 XX 12-AUG-2004 (first entry)
 DE Human secreted protein encoding sequence SEQ ID #506.
 XX Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
 XX cancer; inflammatory; immune; ds; human secreted protein.
 OS Homo sapiens.
 XX MO2004035732-A2.
 XX 29-APR-2004.
 XX 28-AUG-2003; 2003WO-US026780.
 XX 29-AUG-2002; 2002US-0406576P.
 PR 29-AUG-2002; 2002US-0406579P.
 PR 29-AUG-2002; 2002US-0406585P.
 PR 29-AUG-2002; 2002US-0406588P.
 PR 29-AUG-2002; 2002US-0406608P.
 PR 29-AUG-2002; 2002US-0406611P.
 PR 29-AUG-2002; 2002US-0406612P.
 PR 29-AUG-2002; 2002US-0406614P.
 PR 29-AUG-2002; 2002US-0406642P.
 PR 29-AUG-2002; 2002US-0406646P.

QY	529	GlyTyrGlnIleThrAlaMetAlaAenA	gThrIleLeuValAspAsnAenTh	TripAsn	548
Db	1889	GGCTATCAGATTGCTGGCATGGCCAA	CAGGACCACTCTGGTGGATAACA	CACACCTCGAAC	1948
QY	549	AenThrHisIleSerArgValGlyGln	AlaMetAlaSerThrGluGluLy	AlaTyrGlu	568
Db	1949	AACAGCCACATCGCACTGGTCGGAA	GAAGCTATGCTCTCCAATGAAC	CGCGCCCTATAA	2008
QY	569	IleMetArgGluLeuAspValSerTyr	ValLeuValIlePheGlyGlyLeu	ThrGlyTyr	588
Db	2009	ATCATGAGTCCCTTGATGTCGAAT	TANGTTGGTTATTTTCGGAGAG	GATGTTGGCTAT	2068
QY	589	SerSerAspAlleAsnLysPheLeu	TpMetValArgIleGlyGlySer	ThrAspThr	608
Db	2069	TCCGGGACGATATCAACAGTTCCT	CTCGATGGTCAGGATAGCT-----	GAA	2116
QY	609	GlyLysHis-----IleLysGlu	AenAspTyrThrProThrGlyGlu	PheArg	625
Db	2117	GGGGAGCATCCCAAGACATCCGG	AAGGTGACTATTTCACCCAGCAG	GAGAGTTCGGA	2176
QY	626	ValAspArgGluGlySerProVal	LeuLeuAenCysLeuMetTyrLys	MetCysTyrTyr	645
Db	2177	GTAGCAAGCTGGGTCTCCTACTCT	GTTAACTGCCCTATTGTATAAA	TGTCATACTAC	2236
QY	646	ArgPheGlyGlnValTyrThrGlu	AlaLysArgProCysGlyPheAsp	ArgValArgAsn	665
Db	2237	AGATTGGAGAAATGCAGCTAGAT	TTTCGCACTCCCCCAGCGCTT	TGACCGGAACACGTAAT	2296
QY	666	AlaGluIleGlyAenLysAspPhe	GluLeuAspValLeuGluGluGly	TyrThrThrGlu	685
Db	2297	CTGAGATTGGAATAAAGACATTA	ATATTCACGATTTGGAGGAAG	CGCTTTTACATCAG	2356
QY	686	HisTrpLeuValArgIleTyrLys	ValLysAspLeuAspAsrArg	700	
Db	2357	CACCTGGCTTGCAGGATATATAAG	TGAAAGTGAAGCACCCTCGAAC	CAGG	2401

RESULT 10

ADD94789	ADD94789 standard; DNA; 2417 BP.
XX	
XX	
AC	ADD94789;
XX	
XX	
DT	29-JAN-2004 (first entry)
XX	
XX	Drosophila melanogaster STT3 gene sequence.
DE	
XX	
XX	source of immunodominant MHC-associated peptide; STMP; MHC;
KW	major histocompatibility complex; human leukocyte antigen; HLA;
KW	cystatic; immunosuppressive; antitense therapy; gene therapy; cancer;
KW	lung cancer; intestine cancer; sarcoma; prostate cancer;
KW	testicular cancer; breast cancer; melanomas; pancreatic cancer;
KW	haematological cancer; immune response; lymphoid cell proliferation;
KW	autoimmune disease; transplant rejection; STMP-derived peptide;
KW	fruit fly; gene; ds; STT3.
KW	

PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
PT or breast cancer, or for suppressing an immune response in an autoimmune
PT disease.

PS Claim 6; SEQ ID NO 7; 66pp; English.

This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC)-associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytotoxic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestine cancer, sarcomas, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the *Drosophila* melanogaster SPT3 gene which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from GenBank.

Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other; XX SQ

Alignment Scores:

Argument Scores:					
Pred. NO.:	7.24e-211	Length:	2417		
Score:	2177.50	Matches:	415		
Percent Similarity:	72.92%	Conservative:	110		
Best Local Similarity:	57.64%	Mismatches:	144		
Query Match:	58.90%	Indels:	51		
DB:	10	Gaps:	9		

US-10-028-384-12 (1-705) x ADD94789 (1-2417)

QY	16	ThrLeuLeuLysLeuLeuLeuLeuLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgIleu	35
Db	119	AGCCTAATACACCTTCGGCATCTCCTGTAATAGCCTGGCTGGCGGGATTTTCCTTCGCGCTC	178
QY	36	PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg	55
Db	179	TTCGCGGTCATCCGTTTTCGAGTCGATATTCATGAGTTTGATCCGTTGGTTCACATACCGG	238
QY	56	ThrThrArgPheLeuAlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArg	75
Db	239	GGCACCGGCTACATGGTCGACGAATGGTTGTGTACAACTTCCTCAACTGGTTCGACGAGCGC	298
QY	76	AlaATpTyrProLeuGlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThr	95
Db	299	GCATGGTATTCGGCTCGCAGGATTTGGGGGGTACCGTCTATCCCGGCTCGATGATTACG	358
QY	96	SerAlaAlaIleTyrHisIleValLeuHisPhePheHisIleThrIleAspIleArgAsnVal	115
Db	359	TCCGGCGGAATCCATTCGCTGCTGCAGCTACTCAACATACCGGTCATATTCGTGCACATC	418
QY	116	CysValPheLeuAlaProLeuPheSerSerPheThrSerIleValThrTyrIleLeuThr	135
Db	419	TCCGTGTTTCCTTGGGGCGGATCTTCAGTGGCTGACCTCCCATCTCCACCTACCTGTGACC	478
QY	136	LysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValPro	155
Db	479	AAGGAGCTGTGTCGCGGGCGCGCGCTCTTCGCGCGCCAGCTTCATTCGCCATTCGTGCCT	538
QY	156	GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys	175
Db	539	GGCTACATCAGTAGATCGTGGCTGGATCGTAGATTAACGAGGGCATTCGCATATTCGCC	598
QY	176	MetLeuLeuThrTyrTyrMetTrpIleLysAlaValIleThrGlySerIleCysTrpAla	195
Db	599	CTFCAGAGTTCACTACTCTTCGTGGTGGCTCGCTAGTGAAGACTGCATTCGCTGTTCGTGCG	658

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (AB157737-
 CC AB172072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8,52e-211 Length: 2699
 Score: 2177.50 Matches: 415
 Percent Similarity: 72.92% Conservative: 110
 Best Local Similarity: 57.64% Mismatches: 144
 Query Match: 58.90% Indels: 51
 DB: 4 Gaps: 9

US-10-028-384-12 (1-705) x AB113247 (1-2699)

QY 16 ThrLeuLeuLysLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu 35
 DB 128 AGCTAATACCTTCCGATCCCTGCTAATCGCTGCGCGCGAATTTCTCTCGCCTC 187
 QY 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
 DB 188 TTCGCCGTATCATCGTTTCAGATCGATATCCATGAGTTTGATCCGGTTCACATACCGG 247
 QY 56 ThrThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsnTrpPheAspArg 75
 DB 248 GCACCGCGCTACATGTGCGAGATGTTGGTACACTTCTCAACTGCTTCACGAGCGC 307
 QY 76 AlaTrpTyrProLeuGlyArgIleGlyGlyThrIleTyrProGlyLeuMetIleThr 95
 DB 308 GCATGTATCCGCTCGCAGGATTTGGCGGTACCGCTATCCCGCGCTGATGATTACG 367
 QY 96 SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal 115
 DB 368 TCCGGCGGAATCATGGCTGTGCTGCACGTACTCAACATACCGCTCCATATTCGTGACATC 427
 QY 116 CysValPheLeuAlaProLeuPheSerPheThrSerIleValTyrIleLeuThr 135
 DB 428 TGGCTGTTCTGGCGCGGATTTTCAGTGCCTGACCTCCATCTCCACCTACCTGCTGACC 487
 QY 136 LysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValPro 155
 DB 488 AAGGAGCTGTGTCGCGCGCGCGCTCTTCGCGCGCAGCTTCATCGCATCTGCTGCT 547
 QY 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 DB 548 GGTACATCAGTAGGTCGTGCTGATCGTACGATACGATACGAGGGGATTCGCAATATCGCC 607
 QY 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla 195
 DB 608 CTGCAGTTCACTTCTCTGTTGGTGGCTCAGTGAAGACTGATCGTGTCTGGTGG 667
 QY 196 AlaLysCysAlaLeuAlaTyrPheTyrMetValSerTrpGlyTyrValPheLeu 215
 DB 668 GCGCGAGCGCGCTTGTCTCTCTTCTACATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 727
 QY 216 IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
 DB 728 ATCAACCTGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 787
 QY 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
 DB 788 CTGCTGACGAGCTACAGCACTTCTACATCTCTGGGACTGCTGCTGCTGCTGCTGCTGCTGCT 847
 QY 256 PheValGlyPheGluProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275

DB 848 TTCGGTGATTCCAAACCCGATACGACACAGTGAACACATAGCTGCTGGAGTGTGTTGTG 907
 QY 276 LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPhe 295
 DB 908 CTCTTATGCGCGTGGCCACCTTGCCTATTTGCAGTTCGCTGTCGCGCAACGAGTTC 967
 QY 296 GluValLeuPheArgSerValIleSerLeuValGlyPheValLeuLeuThrValGly 314
 DB 968 CGGAAGCTGTTC-----ATCGTCGCGGATTTGCTGGGCTTGGGCTC 1012
 QY 315 -----AlaLeuLeuMetLeuThr-----GlyLysIleSerProTyrThrGlyArgPhe 330
 DB 1013 TTTGTGGCGGTGGTGTCTCACCATGCTGGCGCTTGTGGCCCGCTGGAGTGGAGCTTC 1072
 QY 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleLeuAlaSerValSer 350
 DB 1073 TACTCGTGTGGGATCTAGTGGTACGCCAAGATCCACATCCCATTCATTCGCTGCTG 1132
 QY 351 GluHisGlnProThrThrTrpSerTyrPheAspLeuGlnLeuLeuValPheMet 370
 DB 1133 GAGCATCAGCCACCACTTGGTCTCTCTTCTTCTGATCTGCACATCTCTGGTGGGCC 1192
 QY 371 PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIle 390
 DB 1193 TTCCAGTGGGAGTGTGTACTGTCATCAAGCAGATCAACGACGAGCGCTTTCGTGGTG 1252
 QY 391 MetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuVal 410
 DB 1253 CTGTACGCCATCAGTGGGTTTACTTCGCTGGTGTGATGGTGGTGTGATGTTGACCTC 1312
 QY 411 AlaProValMetSerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMet 430
 DB 1313 ACGCCGTGGTGCATGCTGGCGGAGTGGCTTTTCGGGACTGTTGGATGTTCTCTG 1372
 QY 431 LysAsnLeuAspIleSerArgPro----- 438
 DB 1373 CAAGAGATTGCTTAAGCAATGGGCACGCCAATAAGCGACGCCACCGAAGTGGATGAA 1432
 QY 439 -----AspLysLysSerLys----- 443
 DB 1433 GCTGAGGATTCATTGAGAAAGACGCTGTACGACAAAGCTGGCAAGCTGAAGCATCGT 1492
 QY 444 -----LysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGly 459
 DB 1493 ACTAAGCATGATGCCCGCAGGAGTACT-----GGCGTACGCTCCCAACCTGAAGATAT 1546
 QY 460 MetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThr 479
 DB 1547 GTTATTTGGCGGTCTAATGCTGTTGATGATGTTGCTGCTCCATGCTGACGTTGGTACC 1606
 QY 480 SerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGly---AspGlySer 498
 DB 1607 AGCAATGCTACTCCAGTCCCTCCATGCTTGGCTTTCACCAACAGTCAAGATGGATCC 1666
 QY 499 ArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisSerThrProGlu 518
 DB 1667 CGCAACATTTAGACGATTTTCAGAGGCTTACTACTGGCTTTTCGCGAGACATCGCCGAT 1726
 QY 519 AspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg 538
 DB 1727 GATGCTCGCTTATGCTTGTGGGATTCAGGATACGATACGAGGATGGGAATGGCAACAGA 1786
 QY 539 ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla 558
 DB 1787 ACCAGCTAGTGGATTAATACGTGGAACAATAGTACATAGCGGTGGTGGCAAGGCA 1846
 QY 559 MetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTrpVal 578
 DB 1847 ATGCTTCAACCGGAGGAGTCTCTACGAAATTTATGACATCTCTGACGCTGACTACGTT 1906
 QY 579 LeuValIlePheGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTrp 598
 DB 1907 TTGGTATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGATATCAACAGTTCTCTGTGG 1966

QY 599 MetValArgIleGlySerThrAspThrGlyLysHis-----IleLysGluAsn 615
 Db 1967 ATGGTCCGAATTGCT-----GAGGAGAGCATCCCAAGGACATTAAAGGAAGC 2014
 QY 616 AspTyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeu 635
 Db 2015 GATTACTTTTCCGACCGCGGTGAATTACGGGTAGATCCGAGGTGCTCCGGCCCTGCTC 2074
 QY 636 AsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLys 655
 Db 2075 AACTGCCATTATACAAATTAAAGTACTACAGATTCCGGGAATTGAAAGTTGGACTACAGA 2134
 QY 656 ArgProProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeu 675
 Db 2135 GGTCCATCTGGATATGATCGCACACGTAAACCCGCTCATTTGGGAATAAGGACTTCGATCTG 2194
 QY 676 AspValLeuGluGluGlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLys 695
 Db 2195 ACCTACCTGGAGAGCGCTACACACAGAACTGGCTTGTTCGCATCTATAGGTGAAG 2254
 RESULT 12
 ID ABL13246/c
 XX
 AC ABL13246;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO2001/1042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PP 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR P-PSDB; ABB69143.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 34220; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;
 Alignment Scores: 7.9e-203 Length: 4922
 Pred. No.:

Score: 2103.00 Matches: 415
 Percent Similarity: 67.97% Conservative: 113
 Best Local Similarity: 53.34% Mismatches: 140
 Query Match: 56.88% Indels: 112
 DB: 4 Gaps: 11
 US-10-028-384-12 (1-705) x ABL13246 (1-4922)
 QY 16 ThrLeuLeuLysLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu 35
 Db 3795 AGCCTAAATCACCTCCCATCTCTGCTATATCCCTGCGTGGCCGAGATTTCTCTCGCTC 3736
 QY 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
 Db 3735 TTCGCGGTTCATCGTTTCGAGTCGATTATCCATGAGTTGATCGGTGGTTCACTACCGG 3676
 QY 56 ThrThrArgPheLeuAlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArg 75
 Db 3675 GCCACCGCTACATGTCGAGAATGGTTGGTACAACTTCTCACTGGTTCCGAGCGC 3616
 QY 76 AlaTyrTyrProLeuGlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThr 95
 Db 3615 GCATGGTATCCCTCGGCGAGATTGTGGCGGTACCGTCTATCCGGCGCTCATGATTACG 3556
 QY 96 SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal 115
 Db 3555 TCCGCGGAATCCATTGGCTGTGCACGTACTCAACATACCGTCCATATTCGTGACATC 3496
 QY 116 CysValPheLeuAlaProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThr 135
 Db 3495 TCGGTGTTCTGGCCCGCATCTTCGTGGCTGACCTCCATCTCCACCTACTCTGCTGACC 3436
 QY 136 LysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValPro 155
 Db 3435 AAGGAGCTGTGTCGCGCGCGCTCTTCGCGCGCAGCTTCATCGCATCGTGCCT 3376
 QY 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 Db 3375 GGTACATCAGTAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3316
 QY 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla 195
 Db 3315 CTGCAGTTTCACTACTTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3256
 QY 196 AlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeu 215
 Db 3255 GCGCAGCGCTTGTCTTACTTCTACATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3196
 QY 216 IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
 Db 3195 ATCAACTGATACCTCCCTGACCTCTCTGCTACTGCTCATTTGGGCGAGTACTCGCGCGT 3136
 QY 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
 Db 3135 CTCTCAGCAGTACAGCACCTTCTACATCTCTGGGACTGCTGCTCTCTCATGAGATCCC 3076
 QY 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
 Db 3075 TTCTGGGATTCACACGATACGACCATGAGTGAACATGGCTGGCTGGCTGGAGTGTGTG 3016
 QY 276 LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPhe 295
 Db 3015 CTCTTATGGCGGTGGCCACCTTGGCCATTTCGAGTCCGTGCTGCGCGCAACAGATTC 2956
 QY 296 GluValLeuPheArgSerValIleSerLeuValGlyPheValLeuThrValGly--- 314
 Db 2955 CGGAAGCTGTTCT-----ATCTCGCGGAGATTCTGCTGGCGCTTGGCGTTC 2911
 QY 315 -----AlaLeuLeuMetLeuThr-----GlyLysIleSerProTyrThrGlyArgPhe 330
 Db 2910 TTTGTGGCGTGTGGTGTCTCACCATGCTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTTC 2851
 QY 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleAlaSerValSer 350

QY 611 HisIleLysGluAsnAspTyrThrProThrGlyGluPheArgValAspArgGluGly 630
 DB 1891 GAAATCAAGAAAGAGACTACTTACTACCCAGGAGATATAAGTGGATGAAGTCCA 1950
 QY 631 SerProValLeuLeuAsnCysLeuMetTyrIleYsMetCysTyrTyrArgPheGlyGlnVal 650
 DB 1951 TCATCGCAATGAAGAATCTTTCATGATATAAGTTATCGTATCATAGATTCACCTGAATTG 2010
 QY 651 TyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaGluIle---Gly 669
 DB 2011 TTTGGAGGTAGA-----GATGGTGTGATAGATTGAAACCAACAAATCCAGGCC 2061
 QY 670 AsnLysAspPheGluLeuAspValLeuGluGlyTyrThrThrGluHisTyrPheLeuVal 689
 DB 2062 AATGAAGTACCGAAATGAATGTTGTTCAAGAAGCCCTTCATACAGAAATTCGATTGTG 2121
 QY 690 ArgIleTyrIleValLysAspLeuAspAsnArgGly 701
 DB 2122 AGAATTACAAAGTTAAAGATTGGATTAATGTTGGT 2157

RESULT 14

ID ABT20789 standard; DNA; 2232 BP.
 XX AC ABT20789;

XX DT 16-APR-2003 (first entry)

XX DE Aspergillus fumigatus essential gene #3147.

XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 XX KW cancer; contamination; biofilm; antibody; immune response; ds.

XX OS Aspergillus fumigatus.

XX PN WO200286090-A2.

XX PD 31-OCT-2002.

XX PF 23-APR-2002; 2002WO-US013142.

XX PR 23-APR-2001; 2001US-0285697P.

XX PR 27-APR-2001; 2001US-0287066P.

XX PR 05-JUN-2001; 2001US-0295890P.

XX PR 09-JUL-2001; 2001US-0303899P.

XX PR 31-AUG-2001; 2001US-0316362P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

XX DR WPI; 2003-093124/08.

XX PT New purified or isolated nucleic acids of essential genes of Aspergillus
 XX FT fumigatus, useful for treating or preventing infections by A. fumigatus,
 XX PS or for treating a non-infectious disease in a subject e.g. cancer.

XX PS Disclosure; Page; 175pp; English.

XX CC The invention relates to novel purified or isolated nucleic acids of
 XX CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 XX CC the invention are used to treat or prevent infections by a pathogenic
 XX CC organism such as A. fumigatus, to treat a non-infectious disease in a
 XX CC subject (e.g. cancer), to prevent or contain contamination of an object
 XX CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 XX CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 XX CC expressing recombinant protein for characterisation, screening or
 XX CC therapeutic use, as markers for host tissues in which the pathogenic
 XX CC organisms invade or reside, for comparing with the DNA sequence of A.
 XX CC fumigatus to identify duplicated genes or paralogues having the same or
 XX CC similar biochemical activity and/or function, for comparing with DNA
 XX CC sequences of other related or distant pathogenic organisms to identify
 XX CC potential orthologous essential or virulence genes, for selecting and

CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organisms invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX

Sequence 2232 BP; 473 A; 605 C; 545 G; 609 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7.15e-194 Length: 2232
 Score: 2010.00 Matches: 374
 Percent Similarity: 70.62% Conservative: 126
 Best Local Similarity: 52.82% Mismatches: 170
 Query Match: 54.37% Indels: 38
 DB: 8 Gaps: 8

US-10-028-384-12 (1-705) x ABT20789 (1-2232)

QY 17 LeuLeuLysLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPhe 36
 DB 61 CTTCTGGCGATTATATCTCTGTACCATTCAGCAGCAGCAGTTCGAGTCGATCTTTC 120
 QY 37 AlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThr 56
 DB 121 AGCGTATTCGGCTTCGAGAGTATCATCCAGAACTGACCCGCTTCACTCCAGCA 180
 QY 57 ThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAla 76
 DB 181 ACMAAATACCTTAGTACAGATGTTCTATAGCTTTGGGATTGTTTATGACGGGAACA 240
 QY 77 TrpTyrProLeuGlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSer 96
 DB 241 TGGCATCTCTGGGACGCTGTCCCGGTGGCAGCTATATCCCGCTCTCATGGTCACGAGC 300
 QY 97 AlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCys 116
 DB 301 GCGGTGATCTACCATATCTTTCGATTCTTACTATCCCGCTCGATATTCGCAACATCTGC 360
 QY 117 ValPheLeuAlaProLeuPheSerPheThrSerIleValThrTyrLeuLeuThrLys 136
 DB 361 GTCTACTGGCGCCAGGATTCCTCGGCTGACTGCTGCAATGATGTTCTGTCATCC 420
 QY 137 GluLeuLysAspAla---GlyAlaGlyLeuLeuAlaAlaMetIleAlaValPro 155
 DB 421 GAGATGCTCTTCGCCATCTGCAGGTCTTCTTGACAGCAGCTTTCATGGGAATCGCCCT 480
 QY 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 DB 481 GGTTACATCTCCCGATCAGTTGCTGGAGCTACGATACGAGCGGATTCGTCATCTTCTG 540
 QY 176 MetLeuLeuThrTyrTyrMetTyrIleLysAlaValLysThrGlySerIleCysTyrPala 195
 DB 541 CTTGTGTTCATCTCTTCTATGATCAGGCTGTCAAAATGGTCTATCATGTGGGA 600
 QY 196 AlalysCysAlaLeuAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeu 215
 DB 601 GCGGTGACCGCACTATTCTACGCTACATGGTTCGGCAGGGGTGGTATGTTCTTCAAT 660
 QY 216 IleAsnLeuLeuProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
 DB 661 ACGAACCTGATCCCGCTGCAGCTTTTGTCTCTTGTGCTGTCATGGGTAGATACAGACTCGC 720
 QY 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
 DB 721 ATTACATTAGCTATACCATGTTATGCTGGGACTTTGGTATGTCATGCAGATTCCTCC 780

QY 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
 DB 781 TTGGTGGGATTTTGGCTATCCGAAACAGCAGCACATGTCGGCTTGGGTGTCTTCGGC 840
 QY 276 LeuGlyGlnHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPhe 295
 DB 841 CTGCTTACGCTTGGGCTTCGGCAGTTTGTCCGACGTTCTCCACGACGAGTTC 900
 QY 296 GluValLeuPheArgSerValile-----SerLeuValGlyPheVal 309
 DB 901 CAGACACTTTCAGCCGCAATGCTTCCATCACCCTTCGCTCGTTCGTTGGCTAGTT 960
 QY 310 LeuLeuThrValGlyAlaLeuLeuMetLeuThrGlyLysLeuSerProTyrPheGlyArg 329
 DB 961 GTTCTGACTGTG-----ACGGAGTGAFTCGCTCTTCGATCGAACTTCCTCATCTGG 1002
 QY 330 PheTyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleIleAlaSerVal 349
 DB 1003 TTCTACTCTTGTGGGACTGGCTATGCCAAATCCACATTCCTCATCTTCCTCAGTC 1062
 QY 350 SerGluHisGlnProThrThrTyrSerSerTyrTyrPheAspLeuGlnLeuValPhe 369
 DB 1063 TCGGAACACACAGCCAGCTTGGCCAGCTTCTTTCGATCTGAACCTTCCTCATCTGG 1122
 QY 370 MetPheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIle 389
 DB 1123 CTTTTCCTCCGCGAGGTCTACATGCTTCTCGTACCTCAAGGACGAGCATGCTTCGTC 1182
 QY 390 IleMetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuVal 409
 DB 1183 ATTATCTACTCGCTTTCGCGACTTCTCGCGGTGTATGGTCCGACTAATGCTGACC 1242
 QY 410 LeuAlaProValMetSerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyr 429
 DB 1243 TTGAACCTTATGTGTGTCTCCGCTGCTCGCGCTGTCTGCTCATCTTCGACACTAT 1302
 QY 430 MetLysAsnLeuAspIleSerArgProAspLysSerLysLysGlnGlnAspSer--- 448
 DB 1303 ATGGCGACT--ACCTCCCGACACACCGGCTCTGAAGCGAAACCGAATGAAGACTCGTCT 1359
 QY 449 -----ThrTyrProIleLys 453
 DB 1360 TCACAACTCTTCGCTCAGTTCGGAAGCCAAATGTGAATCACCTCCCATGTTTCTTAAG 1419
 QY 454 IleGluValAlaSerGlyMetIleLeuValMetAlaPheLeuIleThrTyrThrPhe 473
 DB 1420 ATTATAGTACGGCTGTCTGTGCTC-----TACTCTCTCTGTGTTCGCG 1467
 QY 474 HisSerThrTrpValThrSerGluAlaTyrSerProSerIleValLeuSerAlaArg 493
 DB 1468 CACTGCACCTGGGTATACATCAATGATGATCTCTCTCTTCGCTTCTGCTGCTGCTGCT 1527
 QY 494 GlyGlyAspGlySerArgIleIlePheAspAspPheAspGluAlaTyrTyrTrpLeuArg 513
 DB 1528 ATGCTTGACGAAGCAATACATCATGATGATGATGATGATGATGATGATGATGATGAT 1587
 QY 514 HisAsnThrProGluAspAlaLysValMetSerTyrTrpAspTyrGlyTyrGlnIleThr 533
 DB 1588 CAGAATACTCTCAGAACGCCAAATCATGTCATGTTGGGATATGGGTATCAAAATCGGT 1647
 QY 534 AlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTyrAsnAsnHisIleSer 553
 DB 1648 GGCATGCGGACCGCCCACTTGGTTGACACACACCTGGAACACACCATATTCCT 1707
 QY 554 ArgValGlyGlnAlaMetAlaSerThrGluGlyLysAlaTyrGluIleMetArgLeu 573
 DB 1708 ACGGTTGGTAAAGCGATGAGCTCACGCGAGGAAGTCAGTACCTCCCATCTCCGCGCAGCAT 1767
 QY 574 AspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIle 593
 DB 1768 GATGTCGATACGTCGTGGGTGCTCGGTGGTCTCTAGGTATTCGCGCATGACATT 1827
 QY 594 AsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHisIleLys 613

DB 1828 AACAAATTTCTATGATGTCCTGATCGCCGAAGGTATC---TGGCCCGATGAGTTAAA 1884
 QY 614 GluAsnAspTyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerProVal 633
 DB 1885 GAGCGGGACTTCTTCTTACTGACCGGGTGAATATCGTTCGACGATGAGGACCCCACT 1944
 QY 634 LeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGlu 653
 DB 1945 ATGCGCAACAGCTTGTGATATAAATGCTTATTACATTTCAACTCTCTTCCCGTCG 2004
 QY 654 AlaLysArgProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPhe 673
 DB 2005 GCCCAA-----GCTGTGCGACCGCTCGTGGTCAAAACTTCCACAGAGGCCCT 2055
 QY 674 GluLeuAspValLeuGluGlyTyrThrThrGluHisTrpLeuValArgIleTyrLys 693
 DB 2056 CAGCTCTCTACACTCGAAGAGCTTTCACGAGCGAGAACTGATCATCTGTTATCTAAG 2115
 QY 694 ValLysAspLeuAspAsnArgGly 701
 DB 2115 GTCAAGGATCTTGACACCTTGGC 2139
 RESULT 15
 AD36900
 ID ADB69900 standard; DNA; 2466 BP.
 XX ADB69900;
 AC
 XX 04-DEC-2003 (first entry)
 DT
 DE C. neoformans open reading frame SEQ ID NO:2305.
 DE
 XX ds; gene; fungicide; gene therapy; infection.
 KW
 XX C. neoformans open reading frame SEQ ID NO:2305.
 OS
 XX Cryptococcus neoformans.
 PN
 XX WO2003052076-A2.
 PD
 XX 26-JUN-2003.
 PD
 XX 17-DEC-2002; 2002WO-US040225.
 PF
 XX 17-DEC-2001; 2001US-0341261P.
 PR
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX Zamudio C, Eroshkin AM;
 PI
 XX WPI; 2003-533017/50.
 DR
 XX P-FSDB; ADB70261.
 DR
 XX New nucleic acid, useful for preparing a composition for treating an infection caused by Cryptococcus neoformans.
 PS
 XX Claim 2; SEQ ID NO 2305; 136pp; English.
 CC
 XX The invention relates to a novel purified or isolated Cryptococcus neoformans nucleic acid molecule comprising a sequence encoding a polypeptide comprising a sequence not given in the specification. A polynucleotide of the invention has fungicide activity, and may have a use in gene therapy. The nucleic acid is useful for preparing a composition for treating an infection caused by Cryptococcus neoformans. The present sequence represents a C. neoformans sequence of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 CC
 XX
 SQ Sequence 2466 BP; 508 A; 678 C; 580 G; 700 T; 0 U; 0 Other;

Alignment Scores: 1.15e-185 Length: 2466
 Pred. No.: 1930.00 Matches: 373
 Score:

us-10-028-384-12.rng

Thu Dec 16 16:25:00 2004

QY 671 sAspPheGluLeuAspValLeuGluGluGlyTyrThrThrGluHisTrpLeuValArgI 691
 Db 2271 CAGTGTACTCTTGATACCTTGTGACGAAGCGTTCCACATCCGAAATTTGGATCGTCAGGAT 2330
 QY 691 eTyrLysValLysAspLeuAspAsnArgGly 701
 Db 2331 CTACAGGTCAGAGGAGGATCCCATTTGA 2361

Search completed: December 14, 2004, 20:07:11
 Job time : 905.511 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	1929	52.2	2157	4	US-09-614-221A-318
2	1298	35.1	1386	4	US-09-270-767-11648
3	965	26.1	914	4	US-09-270-767-12856
4	930	25.2	900	4	US-09-270-767-12856
5	805	21.8	487	3	US-09-270-767-12331
6	776.5	21.0	867	4	US-09-385-982-213
7	774.5	20.9	1660	4	US-09-248-796A-3089
8	774.5	20.9	1660	4	US-09-270-767-158
9	683	18.5	503	4	US-09-270-767-15440
10	648.5	17.5	507	4	US-09-270-767-27262
C 11	611.5	16.5	1660	4	US-09-513-999C-1965
C 12	611.5	16.5	1660	4	US-09-270-767-158
C 12	611.5	16.5	1660	4	US-09-270-767-15440
					Sequence 118, App
					Sequence 11647, A
					Sequence 12856, A
					Sequence 12331, A
					Sequence 213, App
					Sequence 3089, App
					Sequence 158, App
					Sequence 15440, A
					Sequence 27262, A
					Sequence 1965, App
					Sequence 158, App
					Sequence 15440, A

QY 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
 DB 100 TTTTCAGTCATCAAAATTTGAGTCTATATCCATGAATTCGACCCCTGGTTCAATATAGG 159
 QY 56 ThrThrArgPheLeuAlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArg 75
 DB 160 GCTACCAAAATCTCGTCAACAATTCGTTTACAAAGTTTTCGAACTGGTTTGACACCGT 219
 QY 76 AlaTrpTyrProLeuGlyArgIleIleGlyGlyThrIleTyrProGluMetIleThr 95
 DB 220 ACTGGTACCCCTCGAAGGGTTACTGGAGGACTTTATATCCCTGGTTTGAATGACGACT 279
 QY 96 SerAlaAlaIleTyrHisValLeu---HisPhePheHisIleThrIleAspIleArgAsn 114
 DB 280 AGTCGGTTCATCTGGCAGCCGCTCGCACTGGTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 339
 QY 115 ValCysValPheLeuAlaProLeuPheSerPheThrSerIleValThrTyrLeuLeu 134
 DB 340 GTTTGTGTGCTATTTGGCCCACTATTTCTTGGGCTCACCCTGGGCACTTACGAATTT 399
 QY 135 ThrLysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValVal 154
 DB 400 ACGAAACAGATTAAAGATGCGCAGCGCTGGCTTTGGCTGCTGGTTTATAGCCATTGTC 459
 QY 155 ProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePhe 174
 DB 460 CCCGTTATATATCTAGATCAGTGGCGGGTCTCAGATANTAGGCCATTGCCATTACA 519
 QY 175 CysMetLeuLeuThrTyrTyrMetTrpIleLysAlaValIleThrGlySerIleCysTrp 194
 DB 520 CTATTATGTCATCTTCATGTTTGGATTAAAGCCCAAAAGACTGGCTCTATCATGTCAC 579
 QY 195 AlaAlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPhe 214
 DB 580 GCAACGTGTGCGAGCTTTATCTACTTCTACATGGTGTGCGCTGGGGTGGATACGTGTTC 639
 QY 215 LeuIleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHis 234
 DB 640 ATCACCACATGTCACCTCCATCTCTTTTCTGATTTGATGGCAGATATTTCGTC 699
 QY 235 ArgIleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIle 254
 DB 700 AAACGTGATCTGCTACACCATCTGGTACGCTATTGGAACTGTGTCATCCATCGAGATC 759
 QY 255 SerPheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPhe 274
 DB 760 CCATTTGTCGGTTCTCTACCTATCAGTCTTAACGACCATGCGCCGCTATGGGTGTTTC 819
 QY 275 GlyLeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGln 294
 DB 820 GGTITGATTCAGATTGCGCTTCGGTGACTCTGTAAGGGCCAAATCAGACACGCTAAG 879
 QY 295 PheGluValLeuPheArgSerValIleSerLeuValGlyPheValLeuLeuThrValGly 314
 DB 880 TTTAAAGTCATCATG-----ATGGTTTCTCTG-----TTTTGATCTTGGCTCTGGT 927
 QY 315 AlaLeu-----LeuMetLeuThrGlyLysIleSerProTyrThrGlyArgPhe 330
 DB 928 GTGTCGGACTTCTGCTTGACTATATGGGGTTGATTTGCCCTTGGACTGGTAGATT 987
 QY 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAlaProIleIleAlaSerValSer 350
 DB 988 TATTTCGTTATGGGATACCAACTACGCAAGATCCCATTCCTATCATTTGCTCCCTTCC 1047
 QY 351 GluHisGlnProThrThrTrpSerSerTyrTyrPheAspLeuGlnLeuValPheMet 370
 DB 1048 GAAATCAACCCGTTTCGTCGGCGCTTCTCTTTGATACCCACTTTTGTATCTGGCTA 1107
 QY 371 PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIle 390
 DB 1108 TTCCCGCCCGGTGATTCCTACTATTCCTCGACTTGAAGAGCAGCAGCTTTTGTCTATC 1167
 QY 391 MetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeu 410

DB 1168 GCTTACTCGTTCGTGTGCTGCTACTTTCGCGGTGTTATGGTTAGATTGATGTTGACTTTC 1227
 QY 411 AlaProValMetSerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMet 430
 DB 1228 ACACCAAGTCATCTGTGTGCTCGCGCGCTGCGATTGTCCAAGATATTTGACATCTAC--- 1284
 QY 431 LysAsnLeuAspIleSerArgProAspLysLysSerLysGlnGlnAspSerThrTyr 450
 DB 1285 -----CTGGATTTCAACACAGTGACCGCAA-----TAC 1314
 QY 451 ProIleLysIleGluValAlaSerGlyMetIleLeuVal-----MetAlaPhePhe 467
 DB 1315 GCCATCAAAACCTGGCGCACTACTGCGCAAAATGATGTTTCCGGATCATTCATCTTTAT 1374
 QY 468 LeuIleThrTyrThrPheHisSerThrTrpValThrSerGluAlaTyrSerSerProSer 487
 DB 1375 TTGTATCTTTTCGTTCCATCTTCTACTTTGGTTACAGAACTGCATACCTCTCTCTCTCT 1434
 QY 488 IleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgGlu 507
 DB 1435 GTTGTGTTTGGCATCACAAACCCAGATGGTAAATTTGGCGTTGATCGACGACTTCAGCGAA 1494
 QY 508 AlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrpAsp 527
 DB 1495 GGTACTATTGGTTAAGATGAATCTCTGATGAGGACAGTAAAGTTGCAGCGTCTGGAT 1554
 QY 528 TyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrp 547
 DB 1555 TAGCGTTACCAAAATTTGGTGGCATCGCAGACAGACCACTTTAGTCCGATACACACGCTGG 1614
 QY 548 AsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyr 567
 DB 1615 AACATATCTCATCGCATCGCATCGTTGGTTAAAGCCATGCTTCCCTGAAGAGAAATCTTAC 1674
 QY 569 GluIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGly 587
 DB 1675 GAAATCTTAAAGACATGATGTCGATATTGCTTGTGTCATCTTGTGCTCTAATTTGGG 1734
 QY 588 TyrSerSerAspAspIleAsnLysPheLeuTrpMetValArgIleGlySerThrAsp 607
 DB 1735 TTTGCTGGTGTGATCATCAACAAATCTTGTGGATGATCAGATTAGCGAGGGAATC--- 1791
 QY 608 ThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThrGlyGluPheArgValAsp 627
 DB 1792 TGGCCAGAGAAGATATAAGACGCTGATTTCTATACCCGACAGGAGAAATACAGATAGAT 1851
 QY 628 ArgGluGlySerProValLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPhe 647
 DB 1852 GCAAGGGCTTCTGAGACCATGAGGAATCGCTACTTTTACAGATGCTCTACAAAGATTTC 1911
 QY 648 GlyGlnValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaGlu 667
 DB 1912 CCACAATATTCAATGTTGGGCGCAA-----GCCACTGACACAGTGGCTCAACAAATG 1962
 QY 668 IleGlyAsnLysAspPhe---GluLeuAspValLeuGluGlyTyrThrThrGluHis 686
 DB 1963 ATCACACCATTAGACGCTCCACCATTAGACTACTTCGACGAAGTTTACTTCCGAAAC 2022
 QY 687 TrpLeuValArgIleTyrLysValLysAspLeuAspAsnArgGly 701
 DB 2023 TGGATGGTTAGAAATATATCAATTGAAGAGGATGATGCCAAGT 2067

RESULT 2

US-09-270-767-11648
 : Sequence 11648, Application US/09270767
 : Patent No. 6703491
 : GENERAL INFORMATION:
 : APPLICANT: Homburger et al.
 : TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 : FILE REFERENCE: File Reference: 7326-094
 : CURRENT APPLICATION NUMBER: US/09/270.767
 : CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 11648
 ; LENGTH: 1386
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-11648

Alignment Scores:

Pred. No.: 4,94e-148 Length: 1386
 Score: 1298.00 Matches: 239
 Percent Similarity: 89.54% Conservative: 39
 Best Local Similarity: 76.11% Mismatches: 36
 Query Match: 35.11% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-270-767-11648 (1-1386)

QY 9 LeuSerTyrGluLysGlnAspThrLeuLeuLysLeuLeuLeuSerMetAlaVal 28
 DB 444 CTCACCTGGGACAAACAGGACCACTGGTCAAGCTGGCCATTCTCATCTGGGAGCGGTT 503
 QY 29 LeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValHisGluPhe 48
 DB 504 TTATCATTTGCCACAGCGTTTCTCTGCTGCTGCTGATTCGAAAGCGGTATCCATGAGTTC 563
 QY 49 AspProTyrPheAsnTyrArgThrArgPheLeuAlaGluGluGlyPheTyrLysPhe 68
 DB 564 GATCGGTACTCACTACCGCACCACTGGGTTCTTGGCGGAGGAGGCTTTTACAAGTTC 623
 QY 69 HisAsnThrPheAspArgAlaTyrPheLeuGlyArgIleGlyThrIle 88
 DB 624 CACAACCTGGTTCGATGACCGGCTGTGTATCCCTGGCGGCGATCATCGCGCACCATC 683
 QY 89 TyrProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHisPheHisIle 108
 DB 684 TATCCGGGCTGATGCTCACCTTCGGCGGCTGTACCGCTGTGTGGTGTCTCAATGTG 743
 QY 109 ThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerPheThrSer 128
 DB 744 ACCATCGACATACCGAAGCTGTGGTCTTCTCTGGCGGCTTCTCTCTCTGTGACGACG 803
 QY 129 IleValThrTyrLeuLeuThrLysGluLeuLysAspAlaGlyLeuLeuAlaAla 148
 DB 804 CTGGTGACTACGGCTTCAACAGGAGATACACAGCACTGGAGCTGGATGGTGGCGGCC 863
 QY 149 AlaMetIleAlaValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 168
 DB 864 GCTTTGATATCATCGTTCCCGGATATCTCTCGATCCGTGGCGGATCGTACGACAAT 923
 QY 169 GluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetTrpIleLysAlaValLys 188
 DB 924 GAAGGCATCGCATTTTCTGATGCTCTTCACTACTATTGTGGATCAAGCGGTAAAG 983
 QY 189 ThrGlySerIleCysTrpAlaAlaLysCysAlaLeuAlaTyrPheTyrMetValSerSer 208
 DB 984 ACGGACAGCATCTTTGGTGGCTATGTCGGCATGTCGGCTACTTCTATATGTTCTCTCTG 1043
 QY 209 TrpGlyGlyTyrValPheLeuLeuAsnLeuLeuProLeuHisValLeuValLeuMetLeu 228
 DB 1044 TGGGGTGGCTATGTTCTTCTGATTAACCTAATCCCGCTGCACGTGTGGCGGTGATGATC 1103
 QY 229 ThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrValTyrCysLeuGlyThr 248
 DB 1104 ACCGACGCTTTCTCGCAGAGATCTATACGATACACGATACGATATCTACTGCTCGGCACC 1163
 QY 249 IleLeuSerArgGlnIleSerPheValGlyPheGlnProValLeuSerGluHisMet 268
 DB 1164 ATCTGTGATGATGATGATCTGTTGTGGATTCACACCATCCAGAGCTCCGACACATG 1223
 QY 269 AlaGlyPheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeuArgSer 288
 DB 1224 CTGGCACTGGGAACCTTTGGGCTGTGCCAGATTCACGCTTTCTGCTGCACTATCTGCGCTG 1283

QY 289 LysLeuAsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPhe 308
 DB 1284 GGCATTCCCAAGGATCATTGATTCGCTCTTCAAGACGTTGGTTCCAGTGTTTTACT 1343
 QY 309 ValLeuLeuThrValGlyAlaLeuLeuMetLeuThrGlyLys 322
 DB 1344 GTGGTGTTCGTCGTGGGTACCTGCTCACGCTTACCGGAAA 1385

RESULT 3

US-09-270-767-12856
 ; Sequence 12856, Application US/09270767
 ; Patent No. 8703491
 ; GENERAL INFORMATION:

; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; INVENTION FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 12856
 ; LENGTH: 914
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-12856

Alignment Scores:

Pred. No.: 1,14e-107 Length: 914
 Score: 965.00 Matches: 175
 Percent Similarity: 92.04% Conservative: 10
 Best Local Similarity: 87.06% Mismatches: 16
 Query Match: 26.10% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-270-767-12856 (1-914)

QY 502 PheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLys 521
 DB 2 TTCATGACTTCCGAGGCTACTACTGGCTGCAGATGAACACTCCGAGACGCTCGC 61
 QY 522 ValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeu 541
 DB 62 ATATGTCCTGTGGGACTACGCTACCAAGATACGGCATGGCAATCGGACGATATTA 121
 QY 542 ValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSer 561
 DB 122 GTGATTAACAATCTTGGAAACAACACATATATCCGCGTGGCCAGCGATGGCTCT 181
 QY 562 ThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIle 581
 DB 182 TCGGAGGAGAAAGCTACGAGATATAGGGAACCTGGATGTGGACTACCTTCTCGTATT 241
 QY 582 PheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArg 601
 DB 242 TCGAGGGCTCACTGCTACTCATCGACGATATCAACAGTTCCTGTGGATGGTGGC 301
 QY 602 IleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrTrpThr 621
 DB 302 ATTGGCGGAGCAGGATCGTGTGGCGACATCCGGAAGAGGACTACTATGGGCGAAC 361
 QY 622 GlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLys 641
 DB 362 GGAAGTTCGAGTGGCAAGGAGGCTCACCCACACTGCTCAATTTGTGTATGATCAAG 421
 QY 642 MetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAsp 661
 DB 422 ATGTGCTACTATCGCTTGGCAATATGATACGAGAGGTGGCAAGCCCGGCTACGAT 481
 QY 662 ArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGly 681
 DB 482 CGAGTTCTGTCGCGGAGATCGCAACAAGGACTTTTGAACGTGGATGCTCTGGAGAGCG 541

Qy 682 TyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArgGly 701
 Db 542 TACACACGAGGACACGCTGGTGGCGCACTACAGGTTAAGGATCTCCGATCGTGA 601
 Qy 702 Leu 702
 Db 602 GTC 604
 RESULT 4
 US-09-270-767-12331
 ; Sequence 12331, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 12331
 ; LENGTH: 900
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-12331
 Alignment Scores:
 Pred. No.: 2,09e-103 Length: 900
 Score: 930.00 Matches: 186
 Percent Similarity: 81.32% Conservative: 43
 Best Local Similarity: 64.59% Mismatches: 48
 Query Match: 25.16% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-12 (1-705) x US-09-270-767-12331 (1-900)
 Qy 16 ThrLeuLeuLysLeuLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu 35
 Db 130 AGCTTAATACCTTCGCGCATCTGCTAATCGCGTGGCTGGCGGATTTCTCTCGCGTC 189
 Qy 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
 Db 190 TTCCGCGTCATCGCTTCGAGTCGATTATCCATGAGTTGATCGCTGTTCAACTACGG 249
 Qy 56 ThrThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsnTrpPheAspArg 75
 Db 250 GCCACCGCTACATGGTGGAGATGGTGGTACAACTTCTCAACTGGTTCGACGAGCGC 309
 Qy 76 AlaTrpTyrProLeuGlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThr 95
 Db 310 GCATGGTATCGCTCGGACGATTGTGGCGGTACCGTCTATCCCGGCTGATGATTACG 369
 Qy 96 SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal 115
 Db 370 TCCGCGGATTCATTCGCTGTCGACGACTCAACATACCGGTCATTCGTCGACATC 429
 Qy 116 CysValPheLeuAlaProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThr 135
 Db 430 TGCCTGTCTCTGGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGTCGAC 489
 Qy 136 LysGluLeuLysAspAlaGlyValGlyLeuLeuAlaAlaMetIleAlaValPro 155
 Db 490 AAGAGCTGTGGTCCGCGCGCGCGCTTCGCGCGCGGATTCATCGCATCTGTCCT 549
 Qy 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 Db 550 GGCTACATCAGTAGTCCGTCGCTGATCGTACGATACGAGGCGATTCCTCATTCGCC 609
 Qy 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla 195
 Db 610 CTGAGTTCACCTACCTTCCTGGTGGCGCTCAGTAGAGATCGATCGTGTCTGGTCG 669
 Qy 196 AlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTrpGlyTyrValPheLeu 215

Db 670 GCGCGAGCGCTTGTCTCTACTTCTACATGCTGTCGCGCTGGTGGTGGTCTACGTGTTTCATC 729
 Qy 216 IleAnLeuIleProLeuHisValLeuMetLeuThrGlyArgPheSerHisArg 235
 Db 730 ATCAACCTGATACCCCTGCAGCTTCTGCTACTGCTTATATGGCAGGACTACTCGCGCGT 789
 Qy 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
 Db 790 CTGCTGACCACTACAGCACCTTCTACATCTCGGACTGCTGTCTCTCCATGACATCCCC 849
 Qy 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGly 272
 Db 850 TTCGTGGGATTCACCAACGATACGACGACGATGACACATGCTGGCTGGGA 900
 RESULT 5
 US-09-385-982-213
 ; Sequence 213, Application US/09385982
 ; Patent No. 6262334
 ; GENERAL INFORMATION:
 ; APPLICANT: ENDEGE, WILSON O., ET AL.
 ; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
 ; FILE REFERENCE: CCDNA-260XX
 ; CURRENT APPLICATION NUMBER: US/09/385,982
 ; CURRENT FILING DATE: 1999-08-30
 ; EARLIER APPLICATION NUMBER: 09/328,111
 ; EARLIER FILING DATE: 1999-06-08
 ; EARLIER APPLICATION NUMBER: 60/117,393
 ; EARLIER FILING DATE: 1999-01-27
 ; EARLIER APPLICATION NUMBER: 60/098,639
 ; EARLIER FILING DATE: 1998-08-31
 ; NUMBER OF SEQ ID NOS: 544
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 213
 ; LENGTH: 487
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-385-982-213
 Alignment Scores:
 Pred. No.: 1.33e-88 Length: 487
 Score: 805.00 Matches: 155
 Percent Similarity: 98.73% Conservative: 0
 Best Local Similarity: 98.73% Mismatches: 2
 Query Match: 21.77% Indels: 0
 DB: 3 Gaps: 0
 US-10-028-384-12 (1-705) x US-09-385-982-213 (1-487)
 Qy 242 ThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGlnPro 261
 Db 1 ACTGTTTACTGCTGGGCACTATATCTTCTATGAGATCTCTTTGGGTTTCCAGCT 60
 Qy 262 ValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHisAla 281
 Db 61 GTCCTTTCTATCAGACACATGCGAGCTTTGGGCTCTTTGGTCTCTGCCAGATCCATGCC 120
 Qy 282 PheValAspTyrLeuAspTyrSerLysLeuAsnProGlnGlnPheGluValLeuPheArgSer 301
 Db 121 TTTTGGATTACCTGCGGAGCAAGTTGAATCCACAAATTTGAAGTTCTTTTCCGAGC 180
 Qy 302 ValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuMetLeuThrGly 321
 Db 181 GTCATCTCTCTGGTAGGCTTTGTCTTCTCACCGTGGAGCTCTCTCATGCTGACAGGA 240
 Qy 322 LysIleSerProTyrTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLysAsn 341
 Db 241 AAAATATCTCCCTGAGCGGCGGCTTTCTACTACTGCTGATCCCTCTTATGCTAAGAC 300
 Qy 342 AsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyrTyr 361
 Db 301 AACATCCCCCATCATGCTTCTGTGTCTGAGCATCAGCCCAACCTGGTCTCTCATACTAT 360

QY 362 PheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSerAsn 381
 Db 361 TTGGACCTGAGCTCTCTGCTTTCATGTTCCAGTTGGCTCTATTACTGCTTAGCAAC 420
 QY 382 LeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyr 398
 Db 421 CTGCTGTATGCCGCGATTTTATCATCATGATGATGCTGTGACCAACATGATC 471

RESULT 6

US-09-248-796A-3089
 ; Sequence 3089, Application US/09248796A
 ; Patent No. 6747137

; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
 ; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; PRIOR FILING DATE: 1999-02-12

; PRIOR FILING DATE: 1999-02-12
 ; PRIOR FILING DATE: 1998-02-13

; PRIOR FILING DATE: 1998-02-13
 ; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 3089

; LENGTH: 867
 ; TYPE: DNA

; ORGANISM: Candida albicans
 ; FEATURE:

; NAME/KEY: unsure
 ; LOCATION: (25)

; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unkno

US-09-248-796A-3089

Alignment Scores:

Pred. No.: 1,11e-84 Length: 867
 Score: 776.50 Matches: 147
 Percent Similarity: 76.13% Conservative: 38
 Best Local Similarity: 60.49% Mismatches: 53
 Query Match: 21.00% Indels: 5
 DB: 4 Gaps: 3

US-10-028-384-12 (1-705) x US-09-248-796A-3089 (1-867)

QY 460 MetIleLeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTrpValThr 479
 Db 52 GTTTTACTGACATTTTACATTTTCTACTTTGTTTACATTTGTTTGGGTAAACA 111
 QY 480 SerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArg 499
 Db 112 TCGAATGCTTATTCATCACCATCAGTTGTTTNGCATCCAGAAACCCAGATGGCTCACAA 171
 QY 500 IleIlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAsp 519
 Db 172 CATATCATTTGATATAGAGAGCCTATTACTGTTTAAAGATGAATACACACAGAGAT 231
 QY 520 AlaValValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThr 539
 Db 232 GCCAAGATTATGGCTTGGGTGGGATATGTTTATCAATCGGGGTATGGCTGATAGAACC 291
 QY 540 IleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMet 559
 Db 292 ACATCTGTTTGAACAATACATGGAATAACACACATATTCACATTTGCCATGTTGTTAAGCAATG 351
 QY 560 AlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeu 579
 Db 352 TCTTCCCTCGAGATGTGCTGATGAATTTTGAGACACACAGATGTTGATTATGTGTTA 411
 QY 580 ValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMet 599
 Db 412 GTTATATTTGGAGGTTATTTGGTTATCTGTGTGATGATTTTAACAATTTCTATGGATG 471

QY 600 ValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThr 619
 Db 472 GTAAGAATTGCTGAAGGTATC--TGGCTGATGAATCAAAAGAAAGAGACTACTTTACT 528
 QY 620 ProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMet 639
 Db 529 GACCGGAGAAATATAAGTGGATTAAGATGATCATCTGGCAATGAAGAATTCCTTTGATG 588
 QY 640 TyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGly 659
 Db 589 TATAAGTTATCGTATCATAGATTCACTGAATTTGTTGGAGGTAGA-----CATGGT 639
 QY 660 PheAspArgValArgAsnAlaGluIle--GlyAsnLysAspPheGluLeuAspValLeu 678
 Db 640 GTTGATAGAGTTAGAAACCAAAATCCAGCAATGAATGACCACTGATTTGTTGTT 699
 QY 679 GluGluGlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAsp 698
 Db 700 GAAGAAGCCTTCACATCAGAAATTTGGATTGTGAGAAATTTACAAAGTTTAAAGATTGGAT 759
 QY 699 AsnArgGly 701
 Db 760 AATGTTGGT 768

RESULT 7

US-09-270-767-158

; Sequence 158, Application US/09270767
 ; Patent No. 6703491

; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 158
 ; LENGTH: 1660

; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster

US-09-270-767-158

Alignment Scores:

Pred. No.: 6.08e-84 Length: 1660
 Score: 774.50 Matches: 143
 Percent Similarity: 81.09% Conservative: 20
 Best Local Similarity: 71.14% Mismatches: 31
 Query Match: 20.95% Indels: 7
 DB: 4 Gaps: 2

US-10-028-384-12 (1-705) x US-09-270-767-158 (1-1660)

QY 498 SerArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrPro 517
 Db 788 TCCCGCAACATTTTAGACGATTTTACAGAGGCTTTACTTGGCTTTTCGAGAACACTGCC 847
 QY 518 GluAspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsn 537
 Db 848 GATGATGCTCGGTTATGTTCTTGTGGGATTACGGATACAGATAGCGGGAATGGCAAC 907
 QY 538 ArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGln 557
 Db 908 AGAACGACGCTAGTGGGATTAATACTGTTGAAACAAATAGTCACATGCTTGTGACGTGGACTAC 967
 QY 558 AlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyr 577
 Db 968 GCAATGCTTCAACCGAGGAGAGAGTCTTACAGAAATTTATGACATCTTGTGACGTGGACTAC 1027
 QY 578 ValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeu 597
 Db 1028 GTTTTGTGATCTTTGGCGGTGTGATCGGCTATTCTTGGCGATGATATCAACAGATTCCTG 1087
 QY 598 TrpMetValArgIleGlyGlySerThrAspThrGlyLysHis-----IleLysGlu 614

Db 1088 TGGATGGTCCGAATTGCT-----GAGGAGAGCATCCCAAGGACATTAGGAA 1135
Qy 615 AsnAspTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeu 634
Db 1136 AGCGATTACTTACCGACCGGGTGAATTACGGGTAGATCCGAAGGTGCTCCGGCCCTG 1195
Qy 635 LeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAla 654
Db 1196 CTCAACTGCCCTTATGTACAAATTAAAGCTACTACAGATTCCGGGAATTGAAGTTGGACTAC 1255
Qy 655 LysArgProProGlyPheAspArgValArgAsnAlaGluLeuGlyAsnLysAspPheGlu 674
Db 1256 AGAGTCCCATCTGGATATGTCGACACGTAAACCGCTATTGGGAATAAGGACTTCGAT 1315
Qy 675 LeuAspValLeuGluGluGlyTyrThrGluHisTrpLeuValAlaGlyLeuVal 694
Db 1316 CTGACCTACCTGGAGAGGCGCTACACCAAGACACTGGCTTGTTCGCATCTATAGGGTG 1375
Qy 695 Lys 695
Db 1376 AAG 1378

RESULT 8
US-09-270-767-15440
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Alignment Scores:
Pred. No.: 6,08e-84 Length: 1660
Score: 774.50 Matches: 143
Percent Similarity: 81.09% Conservative: 20
Best Local Similarity: 71.14% Mismatches: 31
Query Match: 20.95% Indels: 7
DB: 4 Gaps: 2

US-10-028-384-12 (1-705) x US-09-270-767-15440 (1-1660)
Qy 498 SerArgileilePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrPro 517
Db 788 TCCCGCAACATTTTACAGCATTTACAGAGAGGCTTACTACTGGCTTTTCGAGAACACTGCC 847
Qy 518 GluAspAlaLysValMetSerTrpTyrAspTyrGlyTyrGlnLeuAlaMetAlaAsn 537
Db 848 GATGATGCTCCGCTTATGCTTGGTGGATTACGATACCATAGCCGGAATGSCAAC 907
Qy 538 ArgThrileLeuValAspAsnAsnThrTrpAsnAsnThrHisileSerArgValGlyGln 557
Db 908 AGAACGACGCTAGTGATATATACGTGAACATATGATACATAGCGCTGGTGGCAAG 967
Qy 558 AlaMetAlaSerThrGluGluLysAlaTyrGluLeuMetArgGluLeuAspValSerTyr 577
Db 968 GCAATGCTTCAACCGAGGAGAGTCTTACGAAATATGACATCTTTGACGTGGACTAC 1027
Qy 578 ValLeuValilePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeu 597
Db 1028 GTTTTGGTATCTTGGCGGTGATCGGCTATCTTGGCGATGATATCAACAGTTCCTG 1087
Qy 598 TrpMetValArgileGlyGlySerThrAspThrGlyLysHis-----ileysGlu 614

Db 1088 TGGATGGTCCGAATTGCT-----GAGGAGAGCATCCCAAGGACATTAGGAA 1135
Qy 615 AsnAspTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeu 634
Db 1136 AGCGATTACTTACCGACCGGGTGAATTACGGGTAGATCCGAAGGTGCTCCGGCCCTG 1195
Qy 635 LeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAla 654
Db 1196 CTCAACTGCCCTTATGTACAAATTAAAGCTACTACAGATTCCGGGAATTGAAGTTGGACTAC 1255
Qy 655 LysArgProProGlyPheAspArgValArgAsnAlaGluLeuGlyAsnLysAspPheGlu 674
Db 1256 AGAGTCCCATCTGGATATGTCGACACGTAAACCGCTATTGGGAATAAGGACTTCGAT 1315
Qy 675 LeuAspValLeuGluGluGlyTyrThrGluHisTrpLeuValAlaGlyLeuVal 694
Db 1316 CTGACCTACCTGGAGAGGCGCTACACCAAGACACTGGCTTGTTCGCATCTATAGGGTG 1375
Qy 695 Lys 695
Db 1376 AAG 1378

RESULT 9
US-09-270-767-27262
; Sequence 27262, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 27262
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-27262

Alignment Scores:
Pred. No.: 1.13e-73 Length: 503
Score: 683.00 Matches: 127
Percent Similarity: 86.23% Conservative: 17
Best Local Similarity: 76.05% Mismatches: 23
Query Match: 18.47% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-270-767-27262 (1-503)
Qy 156 GlyTyrileSerArgSerValAlaGlySerTyrAspAsnGluGlyileAlailePheCys 175
Db 2 GGGTATATCTTCGATCCGTCGGGGATCGTAGCAATGAAGGCATGCCATTTCTGC 61
Qy 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerileCysTrpAla 195
Db 62 ATGCTCTTCACTACTATTGTGGATCAAGGGGGTAAAGACGGGCACGATCTTTGGTCG 121
Qy 196 AlalysCysAlaLeuAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeu 215
Db 122 GCTATGTCGCATTTGGCTTACTTCTATATGGTCTCTCTCGGGGTGGCTATGCTTCTCG 181
Qy 216 IleAsnLeuileProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
Db 182 ATTAACCTAATCCCGTGCACGTGCTGGCGCTGATGATCACCGGACGTTTTCGACAGG 241
Qy 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrileLeuSerArgGlnileSer 255
Db 242 ATCTACATACATACACGACGCTTACTGCTCGGACCATCTTCTGTCGATGACATCTCG 301
Qy 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
Db 302 TTTGTGGGATTCACCAACCCATCCAGAGCTCCGAAACACATGCTGCGACTGGGAACCTTGGC 361

QY 276 LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPhe 295
DB 362 CTGTGCCAGATTCAGCTTTTCCTCGACTATCTGGCTCGCGCATCCCAAGGATCACTTC 421
QY 296 GluValLeuPheArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAla 315
DB 422 GATCTGCTCTTCAAGACGTTGGTTCCAGTGTTTTCAGTGTGGTGTTGCTGCGTGGGTACC 481
QY 316 LeuLeuMetLeuThrGlyLys 322
DB 482 CTGCTCACGCTTACCGGAAA 502

RESULT 10

US-09-513-999C-1965
; Sequence 1965, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1965
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 122..505
US-09-513-999C-1965

Alignment Scores:

Pred. No.: 1,87e-69 Length: 507
Score: 648.50 Matches: 120
Percent Similarity: 81.29% Conservative: 19
Best Local Similarity: 70.18% Mismatches: 25
Query Match: 17.54% Indels: 7
DB: 4 Gaps: 2

US-10-028-384-12 (1-705) x US-09-513-999C-1965 (1-507)

QY 483 TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
DB 2 TACTCTAGTCCAAAGTGTAGTCTCGCTCATACATCATGATGCGCACCAAGAAATATCTTA 61
QY 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysVal 522
DB 62 GATGATTTTACAGAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACAGTA 121
QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
DB 122 ATGCTTGGTGGATTAATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGGTG 181
QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
DB 182 GATATAACACCTGGAAATACAGCCACATAGCAGCTGCTGGGAAAGCTATGCTTCTAAT 241
QY 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
DB 242 GAACACAGCAGCCTATAAATCATGAGGACTCTAGATGATGATTGTTTGGTATTTT 301
QY 583 GlyGlyLeuThrGlyTyrSerAspAspIleAsnLysPheLeuTrpMetValArgIle 602
DB 302 CGAGGGGTTATGGCTATCTCGTGATGATCAACAAATTTCTCTGGATGGTAGATA 361

QY 603 GlyGlySerThrAspThrGlyLysHis-----IleLysGluAsnAspTyrTyrThr 619
DB 362 GCT-----GAAGGAGAACATCCCAAGACATTCGGGAAGTGACTATTTTACC 409
QY 620 ProThrGlyGluPheArgValAspArgGlySerProValLeuLeuAsnCysLeuMet 639
DB 410 CCACAGGAGAAATTCGCTGTAGACAAAGAGGATCCCTACTTTGTGTAATTGCCTTATG 469
QY 640 TyrLysMetCysTyrTyrArgPheGlyGlnVal 650
DB 470 TATAAATGTCATACACTACAGATTGGAGAAATG 502

RESULT 11

US-09-270-767-158/c
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

Alignment Scores:

Pred. No.: 4,97e-64 Length: 1660
Score: 611.50 Matches: 115
Percent Similarity: 79.64% Conservative: 18
Best Local Similarity: 68.86% Mismatches: 27
Query Match: 16.54% Indels: 7
DB: 4 Gaps: 2

US-10-028-384-12 (1-705) x US-09-270-767-158 (1-1660)

QY 532 IleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHis 551
DB 796 GTTCGGGAATGGCAACAGACAGCGTAGTGGTAATTAATACGTGGGAACAAATAGTCAC 737
QY 552 IleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArg 571
DB 736 ATAGCGCTGGTTGGCAAGGCAATGTCTTCAACCGAGGAGAGCTCTACGAATTTATGACA 677
QY 572 GluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAsp 591
DB 676 TCTCTTACGCTGAGTACGCTTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCCGCGCAT 617
QY 592 AspIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHis 611
DB 616 GATATCAACAAGTTCTCTGTGGATGGTCCGAATTGCC-----GAGGGAGAGCAT 569
QY 612 -----IleLysGluAsnAspTyrTyrThrProThrGlyGluPheArgValAspArg 628
DB 568 CCCAAGGACATTAAGGAAGCGGATTACTTTTACCGCGCGGTGAATTCAGGTTAGATGCC 509
QY 629 GluGlySerProValLeuLeuAsnLysLeuMetTyrLysMetCysTyrTyrArgPheGly 648
DB 508 GAAGGTCCTCGGCGCTCTCACTGCCTTATGTACAAATTAAGCTACTACAGATTCGGG 449
QY 649 GlnValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaGluIle 668
DB 448 GAATTAAGTGGTACAGAGGCCCATCTGGATATGATCGCACAGTACGCCGCTCAT 389
QY 669 GlyAsnLysAspPheGluLeuAspValLeuGluGlyTyrThrThrGluHisTrpLeu 688
DB 388 GGAATAAGGACTTTGATCTTGACCTACCTGAGGAGGCCCTACACACAGACACATGGCTT 329
QY 689 ValArgIleTyrLysValLys 695

```

Db          328 GTTCGATCTATAGTGTGAAG 308
|||||
|||||
|||||
|||||
|||||
|||
RESULT 12
US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent NO. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and
; FILE REFERENCE: File Reference: 7326-05
; CURRENT APPLICATION NUMBER: US/09/270,
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 629517
; SOFTWARE: Patentin ver. 2.0
SEQ ID NO 15440
LENGTH: 1660
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-15440

```

```

; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-736-457-1655

Alignment Scores:
Pred. No.: 1.42e-57 Length: 487
Score: 551.00 Matches: 104
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 88.14% Mismatches: 4
Query Match: 14.90% Indels: 10
DB: 4 Gaps: 1

US-10-028-384-12 (1-705) x US-09-736-457-1655 (1-487)

QY 477 TrpValThrSerGlu-----AlaTyrSerSerPro 486
Db 29 TGGATATCTGCAGAAATTCGCCCTTTGAGCGCGCCGCGGACAGTCTACTCTCTCG 88
QY 487 SerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspPheArg 506
Db 89 TCCATTGTACTATCTGCCCGTGGGGATGGCAGTAGGATCATATTTGATGACTCCGA 148
QY 507 GluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrp 526
Db 149 GAAGCATATTTATGGCTCCGTCAATAATCTCCAGAGGATCGAGAGTCAATGCTCTGGTGG 208
QY 527 AspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnIleThr 546
Db 209 GATTATGGCTATCAGATTACAGCTATGGCAACCCGAACAATTTTATGTGACCAATAACACA 268
QY 547 TrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAla 566
Db 269 TGGAAATAATACCCATATTTCTCGAGTAGGCGCAGGCAATGCGCTCCACAGAGGAAAGGCC 328
QY 567 TyrGlnIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyCly 584
Db 329 TATGAGATCATGAGGGAGCTCGAATGTCAGCTATGTGCTGGTCAATTTTGGAGGA 382

Search completed: December 15, 2004, 12:00:08
Job time : 165.129 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 15, 2004, 06:02:06 ; Search time 887.605 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 3697

Sequence: 1 MTKFGFLRLSEKQDTLLKL.....HMLVRIYKVDLNRGLSRT 705

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh
-Q/cgn2_1/USPTO_spoil/US0028384/runat_14122004_131523_8140/app_query.fasta_1.3740
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MOB=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10028384 @CIGN 1.1 2577 @runat_14122004_131523_8140
-NCPUS=6 -ICPU=3 -NO WAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAIDS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=7 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3697	100.0	2472	15	US-10-171-581-112
2	3697	100.0	2472	15	US-10-028-384-11
3	3697	100.0	2472	15	US-10-172-118-742
4	3697	100.0	2472	16	US-10-342-887-742
5	3657	98.9	3046	18	US-10-417-375-95
6	3633	98.3	3093	18	US-10-417-375-92
7	3633	98.3	3094	15	US-10-028-384-9
8	3597	97.3	5404	18	US-10-417-375-99
9	3597	97.3	5404	18	US-10-417-375-97
10	2258	61.1	2839	18	US-10-425-115-150745
11	2258.5	61.0	2681	16	US-10-425-114-14408
12	2247	60.8	2481	15	US-10-028-384-1
13	2237	60.5	2710	17	US-10-028-384-3
14	2237	60.5	2779	17	US-10-437-963-99904
15	2177.5	58.9	2417	15	US-10-028-384-7
16	2070.5	56.0	2256	15	US-10-032-585-6323
17	2010	54.4	2232	15	US-10-128-714-7139
18	1940.5	52.5	3141	18	US-10-425-115-130787
19	1930	52.2	2466	16	US-10-320-797-2305
20	1929	52.2	2157	9	US-09-801-368-387
21	1929	52.2	2157	18	US-10-793-639-318
22	1920	51.9	2733	15	US-10-028-384-5
23	1919	51.9	3197	16	US-10-424-593-111541
24	1851	50.1	2603	15	US-10-128-714-6139
25	1851	50.1	3969	15	US-10-128-714-139
26	1851	50.1	4603	15	US-10-128-714-5139
27	1722	46.6	1848	15	US-10-128-714-2139
28	1682	45.5	2882	16	US-10-320-797-1305
29	1682	45.5	4738	16	US-10-320-797-305
30	1680	45.4	1969	15	US-10-128-714-1139
31	1651.5	44.7	2244	17	US-10-437-963-48342
32	1336.5	36.2	1728	16	US-10-424-593-122476
33	1241.5	33.6	1209	15	US-10-106-698-330
34	1241.5	33.6	1209	16	US-10-264-237-412
35	1241.5	33.6	1543	10	US-09-974-879-133
36	1241.5	33.6	1543	10	US-09-305-736-133
37	1241.5	33.6	1543	11	US-09-818-683-133
38	1241.5	33.6	1543	11	US-09-818-683-133
39	1241.5	33.6	1543	16	US-10-621-401-133
40	1097.5	29.7	1094	17	US-10-437-963-99902
41	931	25.2	1828	10	US-09-945-527-62
42	875.5	23.7	2660	16	US-10-264-049-630
43	820.5	22.2	1114	16	US-10-296-115-629
44	805	21.8	487	10	US-09-871-161-213
45	802	21.7	452	9	US-09-878-178-1456

ALIGNMENTS

RESULT 1
US-10-171-581-112
; Sequence 112, Application US/10171581
; Publication No. US2003010426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 112
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: L38961

! DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-112

Alignment Scores:

Pred. No.: 0 Length: 2472
Score: 3697.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-12 (1-705) x US-10-171-581-112 (1-2472)

Qy 1 MetThrIysPheGlyPheLeuArgLeuSerTyrGluYsGlnAspThrLeuLeuYsLeu 20
Db 107 ATGACTAAGTTTGGATTTTTCGATTTCTATATGAGACGACGACACTTTTGAGCTT 166
Qy 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGTCATGGCTGCTGATTATCTCTCCACCTGCTGTTGCTGCTGAGA 226
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTTGAAAGTGTATCCAGTAGTTTGGATTCGGTACTTTTAAATTATCGGACTACAGGTTCTG 286
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrPyrProLeu 80
Db 287 GCTCAGGAGGGTATTATAAATCCATAACTGGTTTGATGACCGAGCGCTGGTACCCCTTG 346
Qy 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGAGCAATCATGTGAGGACAAATTTACCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 406
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGTCGTGTGTTCTGCGCC 466
Qy 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CCT 526
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
Db 527 GCAGGGGCTGGGCTTCTGTGCTGCGCATGATTCGTGTAGTTCCCTGGATATATCTCCGA 586
Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGTGGCTGGCT 646
Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGGATCAAGGCGAGTAAAGACTGGTTCCATCTGTGGCGAGCTAAGTGTGCCCTT 706
Qy 201 AlaTyrPheTyrMetValSerSerTrpGlyTyrValPheLeuLeuAsnLeuIlePro 220
Db 707 GCTTATTTCTACATGGTCT 766
Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db 767 CTCACATCTCTGCTGTGATGCTCACAGGCGCGTTTCTCTCACCGGATCTATGTGGCCTAC 826
Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 827 TGTAATCTTACTGCTGCTGGTACTATACCTTTCTAGSCAGATCTCTTTGTGGGTTTCCAG 886
Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 887 CCTCTCTCTTCTATCAGACATCGGAGGCTTTGGGGTCTTTGGTCTCTGCGCAGATCCAT 946
Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnIlePheGluValLeuPheArg 300
Db 947 GCCTTTGTGATTAACCTGCGCAGCAAGTTGATTCACACAAATTTGAAGTTCTTTTCGG 1006

Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db 1007 AGCGTCATCTCTCTGTTAGGCTTTGCTCTTCTACCGTGGGAGCTCTCTCTCATCTGACA 1066
Qy 321 GlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db 1067 GGAATAATATCTCCCTGAGCGGGCGTTTCTACTACTGCTGGATCCCTCTTAAGTAAG 1126
Qy 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
Db 1127 AACAAACATCCCATCATTCGTTCTGTGTGAGCATCAGCCCAACCTGGTCTCTCATAC 1186
Qy 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
Db 1187 TATTTTGGCTGACAGCT 1246
Qy 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
Db 1247 AACCTGTCTGATGCGCGGATTTTATCATCATGATGATGCTGTCACGACATGACTTTTCA 1306
Qy 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
Db 1307 GCTGTAAATGGTGGCTCTTAATGCTAGTGTGGACCTGTTATGAGCATTCCTCTGSCAT 1366
Qy 421 GlyValSerGlnValLeuSerThrTyrMetIysAsnLeuAspIleSerArgProAspLys 440
Db 1367 GGAGTCTCCAGGTGCTGTCACATACATGAAGAATCTGGACATAAGTCGCGCAGACAAG 1426
Qy 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
Db 1427 AAGACAGAGAGCAACAGGATTCACCTACCTATTAAGATTGAGTGGCAAGTGGGATG 1486
Qy 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
Db 1487 ATACTGGTCATGGCTTCTTCTCATCACCTACACCTTTCATTCACCTGGGTGACAGT 1546
Qy 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
Db 1547 GAGGCTACTCTCTCCGCTCTGTTACTATCTGCGCTGGTGGGATGSCAGTGGATC 1606
Qy 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
Db 1607 ATATTGTGATGACTTCGAGAAAGCATATATTGGTCTCGTCATATATCTCAGAGGATGC 1666
Qy 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
Db 1667 AAGTTCATGCTCTGGTGGATTTGGCTATCAGATTACAGCTATGCGCAACCGAACAT 1726
Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1727 TTAGTGGACAATAACATGGAATAATACCCATATTTCTCGAGTAGGCGAGCAATGGCG 1786
Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
Db 1787 TCCACAGAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGTGACCTATGTCTGGTGC 1846
Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
Db 1847 ATTTTGGAGGCTCCTGGGTATTCCTCTGATGATATCAACAAGTTCTTTGGATGGTC 1906
Qy 601 ArgIleGlyLysSerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
Db 1907 CGGATTTGGAGGAGCAGACAGATACAGGCAACATATCAAGGAGATGACTATTATATCCA 1966
Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1967 ACTGGGAGTTCGCTGTGGACCGCTGGAAGTTCCTCAGTGTCTCACTGCTCACTGTAC 2026
Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
Db 2027 AAGATGTGTACTCTGCTTTGGACAGGTTTACACAGAACCAAGCGTCTCTCAGGCTTT 2086
Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680

Db 1607 ATATTCGATGCTCCGAGAGCATATATTGGCTTCGTCAATAATCCACGAGATCGG 1666
Qy 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
Db 1667 AAGGTCATGCTCCGTGGGATTTGGCTATGAGATACAGCTATGCGCAACCGAACAAAT 1726
Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1727 TTATGGACATACACATGATGATATATCCATATTTCTCGATAGGACGAGCATGGCG 1786
Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
Db 1787 TCCACAGAGAAAGGCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTC 1846
Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
Db 1847 ATTTTGGAGGCTCACTGGGTATTTCTCTGATGATATCAACAAGTTCTTTGGATGGTC 1906
Qy 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
Db 1907 CGGATTCGAGGAGCACAGATACAGCAAAACATATCAAGGAGATGACTATTATCTCCA 1966
Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1967 ACTGGGAGTTCGGTGGACCGTGAAGGTTCTCCAGTCTGCTCAACTGCTCATGTAC 2026
Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
Db 2027 AAGATGTGTACTATCGCTTTGGACAGGTTTACAGAGAGCCAGCGCTCTCCAGGCTTT 2086
Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
Db 2087 GACCGTCTCGAATGCTGAGATGGGAATTAAGACTTTGAGCTTGATGCTCTGGAGGAA 2146
Qy 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
Db 2147 GGCTATACCAAGAACATTTGGCTGGTGGTGGATATACAGGATATACAGGTAAGGACCTGGATAATCGA 2206
Qy 701 GlyLeuSerArgThr 705
Db 2207 GGCTTGTCAAGGACA 2221

RESULT 3

US-10-172-118-742
; Sequence 742, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mac, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002219
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-742

Alignment Scores:

Pred. No.: 0 Length: 2472
Score: 3697.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0
US-10-028-384-12 (1-705) x US-10-172-118-742 (1-2472)
Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 107 ATGACTATGATTGGGATTTTTCGATTGTCCTATGAGAGCAGGACACACATTTTGAAGCTT 166
Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCCTGCTCAATGGCTGCTGATATATCTCTCCACTGCTGTTGTTGCTGCTGAGA 226
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTGGAAGGTATATCCATGAGTTTATCCGTTTATTAATATCGACTACAGGTTCTCTG 286
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaIleTyrProLeu 80
Db 287 GCTGAGGAGGGGTTTTATAAATTCATAAATGTTTATGATGATGATGATGATGATGATGAT 346
Qy 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGACGAATCATTTGGAGGAACAATTTACCAGGTTTAAATGATCACCCTCTGCTGCAATCTAC 406
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTTCACATCACCATCGACATCGACATCGGAATGCTGTGTGTTCTTGGCC 466
Qy 121 ProLeuPheSerSerPheThrSerIleValTyrThrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CTTCTCTTCT 526
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
Db 527 GCAGGGGCTGGCTTCTTCTGCTGCTGCCATGATGCTGTAGTTCTCTGATATATCTCCCA 586
Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGTGGCTGGCTCTCTATGATTAATGAGGATGCTCCATCTTTTGCATGCTACTCACTAC 646
Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGGATCAAGGACAGTAAGACATGGTTCCATCTGTTGGGCGAGCTAAGTGTGCCCTT 706
Qy 201 AlaTyrPheTyrMetValSerSerTrpGlyTyrValPheLeuIleAsnLeuIlePro 220
Db 707 GCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTGTCTCTGATCACTAATCTT 766
Qy 221 LeuHisValLeuLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db 767 CTCACGTCCTGCTGCTGATGCTACAGGCGGTTCTCTCACGGATCTATGTGGCTAC 826
Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 827 TGTACTGTTTACTGCTGGGACTATATCTTTCTAGGCGAGATCTCTTTGTGGGTTTCCAG 886
Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 887 CCTGTCTCTTATCATCAGACACATGGCAGGTTTGGGTTCTTGTGCTCTCTGCCAGATCCAT 946
Qy 281 AlaPheValAspTyrLeuArgSerLysLysAsnProGlnPheGluValLeuPheArg 300
Db 947 GCCTTTGTGGATTACTGCGGACGAGATTTGAATCCACAACAAATTTGAAGTCTTTTCCCG 1006
Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db 1007 AGCGTCATCTCTCTGTTAGGCTTTTGTCTCTCACCCTGGGAGCTCTCTCTCATGCTGCTGACA 1066

QY 321 GlyIysIleSerProThrThrGlyArgPheTy-SerLeuLeuAspProSerTyrAlaLys 340
DB 1067 GAAAAATATCTCCCTGACGGGCGTTTCTACTCACTGCTGGATCCCTCTTATGCTAAG 1126
QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrThrPheSerTyr 360
DB 1127 AACAAATCCCATCAATGCTTCTGTCTGAGCATCAGCCCAACACCTGGTCTCTATAC 1186
QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
DB 1187 TATTTTGACCGGAGCTCCCTGCTTCTCATGTTTCCAGTTGGCTCTTATTACTGCTTTAGC 1246
QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
DB 1247 AACCTGCTGATGCCCGGATTTTATCATCATGATGTTGACCATGATGACTTTTCA 1306
QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
DB 1307 GCTGTAAATGGTGTCTAATGCTAATGTTGGACCTGTTAAGCATTTCTCTGGCAT 1366
QY 421 GlyValSerGlnValLeuSerThrTyrMetIleAsnLeuAspIleSerArgProAspLys 440
DB 1367 GGAGTCTCCAGGTGCTGCACATACATGAAGAATCTGGACATAAGTCCGCCAGACAAG 1426
QY 441 LysSerIlySlysGlnGlnAspSerThrTyrProIleIysIleGluValIleAspGlyMet 460
DB 1427 AAGAGCAAGAACCAACAGGATCCACCTACCTTAAAGATTGAAGTGGCAAGTGGGATG 1486
QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrThrPheValThrSer 480
DB 1487 ATACTGGTCAATGGCTTTCTTCTCATCACTACACCTTCTTCAACCTGGGTGACCAT 1546
QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
DB 1547 GAGGCCACTACTTCTCCGTCATTTGTTACTATCTGCGCGTGGTGGGATGGCATAGGATC 1606
QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
DB 1607 ATATTTGATGATCTCCGAGAAGCATATATTGGCTTCGTCAATAATCTCCAGAGGATGG 1666
QY 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
DB 1667 AAGGTCAATGCTCCTGGTGGATATGGCTATCAGATTACAGCTATGGCAACCCGAAAT 1726
QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
DB 1727 TTAGTGGACAATAACATGAAATAATACCATATTTCTCGAGTAGGGCAGCAATGGG 1786
QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
DB 1787 TCCACAGAGGAAAGCCATAGATCATGAGGAGCTCGATGCTAGCTATGCTGGTGC 1846
QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
DB 1847 ATTTTGGAGCCCTCACTGGGTATTCCTCTCATGATATCAACAAGTTCTTTGGATGGTC 1906
QY 601 ArgIleGlyGlySerThrAspThrGlyIleHisIleIleGluAsnAspTyrThrPro 620
DB 1907 CGGATTGGAGGAGCAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCCA 1966
QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
DB 1967 ACTGGGAGTTCCGTGGACCGGTGAAGTTCTCCAGTGTCTCAACTGCCTCATGTAC 2026
QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
DB 2027 AAGATGTTTACTATCGCTTTTGGACAGGTTTACACAGAACCAAGCGTCTCCAGGCTTT 2086
QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
DB 2087 GACCGTGTCCGAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGCTCTGGAGAA 2146
QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700

DB 2147 GGCTATACCAGAACATTTGGCTGGTCAGGATATACAGGTAAAGACCTGGATATCA 2206
QY 701 GlyLeuSerArgThr 705
DB 2207 GGCTTGTCAAGGACA 2221
RESULT 4
US-10-342-887-742
; Sequence 742, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2599
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-742
Alignment Scores:
Pred. No.: 0 3697.00 Length: 2472
Scores: Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0
US-10-028-384-12 (1-705) x US-10-342-887-742 (1-2472)
QY 1 MetThrIysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeu 20
DB 107 ATGACTAAGTTTGATTTTGGCATTTGCTATGAGACGACGACACACTTTTGAGCTT 166
QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 167 CTCAATCTGTCAATGGCTGTGTTATCTCTCCACTCGTCTGTTGCTGCTGAGA 226
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
DB 227 TTTGAAAGTGTATCCATGAGTTTATCGTACTTTTATTTATTCGAGCTACCAAGTTCTTG 286
QY 61 AlaGluGlyGlyPheTyrLysPheHisAsnTrpPheAspAspAlaTrpTyrProLeu 80
DB 287 GCTGAGGAGGGTTTATAAATCCATACTGGTTTGTATGACCGAGCTGTACCTTTG 346
QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 347 GGACGAATCAATGAGGACAAATTTACCAGGTTTAAATGATCACCTCTCTCTGCAATCTAC 406
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
DB 407 CATGTACTCCATTTTCCACATCCATCCGAAATGCTGTGTGTGTTCTGCTGCTG 466
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140

Qy	1	MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu	20
Db	136	ATGACTAAGTTTGGATTTTTCGGATTGCTCTATGAGAAGCAGGACACACTTTTGAAGCTT	193
Qy	21	LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg	40
Db	196	CTCATTCCTGTCAGTGGCTGCTGATTATTCCTTCTCCACACGTCGTGTTGCTGCTCCTGAGA	255
Qy	41	PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu	60
Db	256	TTTGAAGTGTATCCATGAGTTTGATCCGGTACTTTAATATTCGAGACTACACAGGTTCCTG	315
Qy	61	AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTyrProLeu	80
Db	316	GCTGAGGAGGGTTTTATAAAATCCATAACTGGTGTTCATGACGAGCCTGGTACCCCTTTG	375
Qy	81	GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaalaIleTyr	100
Db	376	GGACGATCATTGGAGGAAACAATTTACCGAGGTTAATGATCACCTCTGCTGCATCTTAC	435
Qy	101	HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla	120
Db	436	CATGTACTCCATTTTTTCCACATCATCACCATCGCAATTCCGAATGCTGTGTGTCTCGGCC	495
Qy	121	ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp	140
Db	496	CCTCTCTCTCTCTCTTCCACCACCATCGTCACGTACCACCTTACCAAGAGCTCAAGGAT	555
Qy	141	AlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValProGlyTyrIleSerArg	160
Db	556	GAGGGGCTGGGCTTCTTGTGCTGCCATGATTGCTGTAGTTCCTGGATATATCTCCCGA	615
Qy	161	SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr	180
Db	616	TCTGTGGCTGGCTCTCATGATAATGAAGGANTGCCATCTTTTGCATGCTACTCACCTAC	675
Qy	181	TyrMetTrpIleLysAlaValIleThrGlySerIleCysTrpAlaAlaLysCysAlaLeu	200
Db	676	TACATGTGATCAGGCGAGTAAAGACTGGTTTCCATCTGTGTGGCAGCTAAGTGTGGCCCTT	735
Qy	201	AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuIlePro	220
Db	736	GCTTATTTCTACATGGTCTCGTCTCATGGGAGGTTATGTGTCTGATCAACTTAATTCCT	795
Qy	221	LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr	240
Db	796	CTCCAGTCTCTGCTGCTGATGCTCACAGGCGGTTTCTCTCACCGGATCTATGTGGCTAC	855
Qy	241	CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln	260
Db	856	TGTACTGTTTACTGCTGGGCACTATACTTCTATCAGATCTCCTTTGTGGGTTTCCAG	915
Qy	261	ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis	280
Db	916	CCTGTCTCTTTCATCAGACACATGGCAGCCTTTTGGGCTCTTTGGTCTCTGCCAGATCCAT	975
Qy	281	AlaPheValAspTyrLeuArgSerIysLeuAsnProGlnPheGluValLeuPheArg	300
Db	976	GCTTTTGTGGATTACTCGGAGAGAGTTGAATCCACAATTTGAAGTCTTCTTTCCGG	1035
Qy	301	SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr	320
Db	1036	AGCGTCATCTCTCTGGTAGGCTTTGTCTCTTCTCACCGTGGAGCTCTCTCATGCTGACA	1095
Qy	321	GlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys	340
Db	1096	GGMAAANAATCTCCCTGGACGGGGGCTTTCTATCTCGCTGCTGGAGNCCCTCTTATGCTAAG	1155
Qy	341	AsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr	360
Db	1156	AACAAATCCCATCTGCTGTCTGTGCTGAGCATCAGCCACCACTGGTCTCTCATAC	1215

Qy	361	TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer	380
Db	1216	TAATTTGACCTGAGGCTCCCTCGTCTTCATGTTTCCAGTTGGCTCTATTACTGCTTTAGC	1275
Qy	381	AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer	400
Db	1276	AACCTGTCGTATGCCCGGATTTTATCATCATGTATGTTGTGACCAAGATGTACTTTTCA	1335
Qy	401	AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle	420
Db	1336	GCTGTAAATGGTGGCTTAATAGCTAGTGTGGCACCTGTTATGTGCAATTCCTCTGGCAAT	1395
Qy	421	GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys	440
Db	1396	GGAGTCTCCAGGTGCTGTCACATACATGAAGAAATCTGGACATAAGTCTGCAGACAAG	1455
Qy	441	LysSerLysGlnGlnAspSerThrTyrProIleLysIleGlnValAlaSerGlyMet	460
Db	1456	AAGAGCAAGAACACAGGAATTCACCTACCTTAAAGAAATGAAGTGGCAAGTGGATG	1515
Qy	461	IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer	480
Db	1516	ATACGTGCTATGGCTTTCTTCTTCATCACCTACACCTTTCAATCAACCTGGGTGACCACT	1575
Qy	481	GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle	500
Db	1576	GAGGCTTACTTCTTCGTCCTCATTTGACTATCTACCTGCTGCGGTGGGATGCGATGAGATC	1635
Qy	501	IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla	520
Db	1636	ATATTTTCATGACTCCAGAGAAGCATATATTGGCTTCGTCAATAATCTCCAGAGGATGCG	1695
Qy	521	LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle	540
Db	1696	AAGGTCAATGCTCGTGGGATTATGGCTATCAGATTACAGCTATGGCAAAACCGAACAAATT	1755
Qy	541	LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla	560
Db	1756	TTAGTGACAAATAACACATGGAAATAATACCATATTTCTCGAGTAGGCGAGGCAATGGCG	1815
Qy	561	SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal	580
Db	1816	TCCACAGAGGAATAAGCTTATGAGATCATGAGGAGCTCGATGTCAAGTATGTCTGTGTC	1875
Qy	581	IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal	600
Db	1876	ATTTTTGGAGGCTCTCACTGGGTATTCCTCTGATGATATCAACAAGATTCCTTTGGATGGTC	1935
Qy	601	ArgIleGlyGlySerThrAspThrGlyLysPheHisIleLysGluAsnAspTyrTyrThrPro	620
Db	1936	CGGATTCGAGGGAGCACAGATACAGGCAACATATCAAGAGAAATGACTATTATATCTCCA	1995
Qy	621	ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr	640
Db	1996	ACTGGGAGTTCGGTGTGGACCGGTGAAGGTTCCTCAAGTGTGCTCAACTGCCCTCATGTAC	2055
Qy	641	LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyLysPhe	660
Db	2056	AAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAAGCCAAAGCGTCTCCAGGCTTT	2115
Qy	661	AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu	680
Db	2116	GACCGTCTCGGAATGCTGAGATTGGGAATTAAGACTTTGAGCTTGATGTCTCTGGAGGAA	2175
Qy	681	GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg	700
Db	2176	GCATATACCACAGNACATTGGCTGGTCAGAGATATACAGGTAAAGGACCTGGATATATCGA	2235
Qy	701	GlyLeuSerArgThr	705
Db	2236	GGCTTTGTCAAGGACA	2250

RESULT 6

US-10-417-375-92
 ; Sequence 92, Application US/10417375
 ; Publication No. US20040219528A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; APPLICANT: Marc Malandro
 ; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
 ; FILE REFERENCE: 529452001600
 ; CURRENT APPLICATION NUMBER: US/10/417,375
 ; CURRENT FILING DATE: 2003-04-15
 ; NUMBER OF SEQ ID NOS: 176
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 92
 ; LENGTH: 3093
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; US-10-417-375-92

Alignment Scores:
 Pred. No.: 0 Length: 3093
 Score: 3633.00 Matches: 694
 Percent Similarity: 98.87% Conservatives: 3
 Best Local Similarity: 98.44% Mismatches: 8
 Query Match: 98.27% Indels: 0
 DB: 18 Gaps: 0

US-10-028-384-12 (1-705) x US-10-417-375-92 (1-3093)

QY	1	MetThrIysPheGlyPheLeuArgLeuSerTyrGluYsGlnAspThrLeuLeuLysLeu	20
DB	112	ATGACTAAGCTTGGAATTTTGGGATTTGCTTATGAGAGCAGACACACTTCTAAAGCTT	171
QY	21	LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg	40
DB	172	CTCATCTGTCGATGGCTGCTGCTGTTATCTTTTCTACTCGCTTTTTCGTGCTGAGA	231
QY	41	PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu	60
DB	232	TTTGAAGTGTATCATGAGTTTGTATCGTACTTTTAAATATCGGACTACCGGTTCTG	291
QY	61	AlaGluGluGlyPheTyrIysPheHisAsnTyrPheAspAspArgAlaTyrProLeu	80
DB	292	GCTGAGGAGGGTTTATAAATTCATTAATCTGTTTGTATGACCGGGCTGGTACCTTTG	351
QY	81	GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr	100
DB	352	GGCCGAATCATTTGGAGAACAAATTTACCAGGTTTAAATGATCATTCTGCTGCAATCTAC	411
QY	101	HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla	120
DB	412	CATGTACTCCATTTCTCCATATCACTAATGACATTCGGATGTCGTGTTTCTCTGGCC	471
QY	121	ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp	140
DB	472	CCACTTTCTCTCTTCCACCCATCCGTTAGTACACCTTACCAGAGAGCTCAAGGAT	531
QY	141	AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg	160
DB	532	GCAGGAGCTGGGCTTCTTGTCTGCTGCATGATTTGCTGATGTTCTCTGGTATATTTCTGA	591
QY	161	SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr	180
DB	592	TCGTAGTGGCTCCCTATGATATGAGGAGTATGCTATCTTTTGTGATCTGCTTACTTAC	651
QY	181	TyrMetTrpIleLysValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu	200
DB	652	TACATGTGGATCAAGGACGAGACTGGTTCATCTATTGGCTGCGCAAGTGTGCGCTC	711
QY	201	AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuIlePro	220
DB	712	GCATTTTCTACATGCTCTTCTCATGGGAGGCTATGCTTCTCTGATCACTTGAATCCT	771
QY	221	LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr	240

DB	772	CTACATGTCCTGGTCTAATGCTCAGAGCGCGTTTTTCTCACCAGATCTACGTAGCTAC	831
QY	241	CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln	260
DB	832	TGTACTGTTTACTGCTGGGACCACTTCTTCTATGAGATTTCCITTTGTTGGTTCCAG	891
QY	261	ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis	280
DB	892	CCCGTCCCTTTCATCAGAACACATGCGAGCTTTGGAGTGTGGTCTCTCTCAGATCCAT	951
QY	281	AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGlnValLeuPheArg	300
DB	952	GCTTTTCGTAGATTACCTGCGAGCAGATTGATCCACAGCAATTCGAGATCTTTTCCGG	1011
QY	301	SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr	320
DB	1012	AGTGTATCTCCCTGGTTGGCTTTGCTCTCTCCTGAGGAGCTCTCTCATGCTAACCA	1071
QY	321	GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys	340
DB	1072	GGAAAAATTTCCCTCGACAGGCGTTTCTACTCTCTGCTGGATCCCTCTTATGCTAAG	1131
QY	341	AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr	360
DB	1132	AATAACATTTCCATTTATGTCATCTGTTTCTGAGCAGCAGGCCCAACCTGGTCTCTCTAC	1191
QY	361	TyrPheAspLeuLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer	380
DB	1192	TATTTTGTATCTACAGCTCTCTGCTTCATGTTCCAGTTGCGCTCTATCTACTGCTTAGC	1251
QY	381	AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer	400
DB	1252	AACTGTCTGATGCTCGGATTTTATCATCATGATGATGCTGTCGACCATGATGATCTTTCA	1311
QY	401	AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle	420
DB	1312	GCTGTAAATGGTGGCTCTAAATGCTGGTATGTCACCTGTTATGTCATCTTTCTGGCATT	1371
QY	421	GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys	440
DB	1372	GCTGTTTCCAGGTCGCTCCACATATATGAAAAATCTGGACATAGTCCGCCAGACAAG	1431
QY	441	LysSerLysLysGlnAspSerThrTyrProIleLysIleGluValAlaLysArgGlyMet	460
DB	1432	AAGACAAAGACACAGCATCTACTTACCTTATTAAGATGAGTGGCGAGTGGGATG	1491
QY	461	IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer	480
DB	1492	ATACTGGTCATGGCTTTTCTCATCACCTACATACAGTTTCTATTCGAGTGGGACAGT	1551
QY	481	GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle	500
DB	1552	GAAAGCTATTCTTCTCCCTCCATTTGACTGCTGCTGCTGGGATGCGCATGAGATC	1611
QY	501	IlePheAspAspPheArgGluAlaTyrTrpTrpLeuArgHisAsnThrProGluAspAla	520
DB	1612	ATTTTGTATGATCTCCGAGAGACGCTATTATTGGTCTCGTCACCAATATCCAGAGATGA	1671
QY	521	LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle	540
DB	1672	AAAGTCATGTCATGGTGGATTTATGGCTACCAATTAATGCAATTCGCAATTCGCAAT	1731
QY	541	LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla	560
DB	1732	TTAGTGGCAATAACACATGGAATAATACCATATTTCTCAGTAGGCGCAGGCAATGGCA	1791
QY	561	SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal	580
DB	1792	TCCACAGAGAGAAAAGCTTATGAATCATGAGGAGCTTGTATGTCAGCTATGCTGCTGC	1851
QY	581	IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal	600

Db 1852 ATTTTGGAGGCTTACTGGGTAATTCCTCGGATGATATCAACAAGTTCTTTGGATGGTC 1911
QY 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrThrPro 620
Db 1912 CGGATTTGGAGGAAGCACAGACAGAGACACATTAAAGGAGATGACTACTACTACTCTCT 1971
QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1972 ACTGGGGAATTCGGTGTGATCGTGGGTTCTCCGGTGTGCTCAACTGCCCTTATGTAC 2031
QY 641 LysMetCysTyrThrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
Db 2032 AAAATGTGTACTACTACCGCTTTGGGCGAGCTTACACAGACCAAGCGCTCCACAGGCTTT 2091
QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysPheGluLeuAspValLeuGluGlu 680
Db 2092 GACCGTGTTCGAATCTCGATGTTGTAATTAAGACATTTGAGCTTGTATGCTGGAGGA 2151
QY 681 GlyThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
Db 2152 GCGTATACACAGACACTGGCTAGTCAGGATATACAAAGGTAAAGGACCTGGATATCGA 2211
QY 701 GlyLeuSerArgThr 705
Db 2212 GCGTTGTCAAGGACA 2226

RESULT 7

US-10-028-384-9
; Sequence 9, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAULT, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SAMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 3094
; LENGTH: 3094
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_008408
; DATABASE ENTRY DATE: 2000-11-01
; RELEVANT RESIDUES: (1)..(3094)
US-10-028-384-9

Alignment Scores:
Pred. No.: 0 Length: 3094
Score: 3633.00 Matches: 694
Percent Similarity: 98.87% Conservative: 3
Best Local Similarity: 98.44% Mismatches: 8
Query Match: 98.27% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-12 (1-705) x US-10-028-384-9 (1-3094)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 112 ATGACTAAGCTTGGATTTTCGGATTGCTATGAGAAGCAGGACACACTTCTAAAGCTT 171
QY 21 LeuIleLeuSerMetAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 172 CTCATCCTGTGATGGCTGCTGTGTATCTTTTCTACTCGTCTTTTGTGCTGCTGAGA 231
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 232 TTGGAAGTGTATCATCCATGAGTTTGATCCGTAATTTAATATCGGACTACCGGTTCTG 291
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTyrProLeu 80

Db 292 GCTGAGGAGGGTTTTATAAATCCATACTGTTTTGATGACGGGCTTGGTACCCCTTTG 351
QY 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100
Db 352 GCGCGAATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCACTTCTCTGCTGCAATCTAC 411
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 412 CATGTACTCCATTTCTTCCATATCACTATTGACATTCGGAATGCTGTGTTTTCTCTGGCC 471
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 472 CCACCTTTCTCCCTTTTCCACCATCTTACGTACCACTTACCAAAAGAGCTCAAGAT 531
QY 141 AlaglyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
Db 532 GCAGAGCTGGCTTCTTGTGCTGCTGATGATGCTGTGCTGGGTATATTCTCGA 591
QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 592 TCTGTAGCTGGCTCTTATGATATGAAGGAATTTGCTATCTTTTGTGATGCTGTTACTTAC 651
QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaLysCysAlaLeu 200
Db 652 TACATGTGGATCAAGGAGTGAAGACTGGTTCCATCTATTGGGCTGCCAAGTGTGCCCTC 711
QY 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIlePro 220
Db 712 GCTTATTTTACATGGTCTCTTTCATGGGAGGCTATGTTCTCTGATCAACTGATTCCT 771
QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db 772 CTACATGCTCTGGTGTAAATGCTGACAGCGCGTTTTTCTCACCGGATCTACATGACCTAC 831
QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 832 TGTACTGTTTACTGCTGGCGGCACCACTTCTTCTATGCAGATTTCTTTTCTGGTTTCCAG 891
QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 892 CCCGTCTCTTTCATCAGAACACATGGCAGCGCTTTGGAGTGTGTTGCTCTCTGTCAGATCCA 951
QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArg 300
Db 952 GCTTTTGTAGATTACCTGGCGCAGCAAGTTGAATCCACAGCAATTCGAATCTTTTCCGG 1011
QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db 1012 AGTGTATCTCCCTGGTGGCTTTGCTCTCTCCTCCTCCTGAGGAGCTCTCTCATGTAACA 1071
QY 321 GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db 1072 GGAAATTTTCTCCCTGGACAGCGGCTTTCTACTCTCTGCTGGATCCCTCTTATGCTAAG 1131
QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
Db 1132 AATAACATTTCCCATTTATGCACTGTTTCTGAGCACCCAGCCACACCTGGTCTTCTCTAC 1191
QY 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
Db 1192 TATTTTGTATCATAGCTCTCTTGTCTTCAATGTTTCCAGTTTGGCTCTTATTTACTGCTTAGC 1251
QY 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
Db 1252 AACCTGTCTGATGCTCGGATTTTATCATCATGATGATGGTGACCCAGCATGACTTTTCA 1311
QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
Db 1312 GCTGTATGTGGCTCTAATGCTGGTATTTGGCACCTGTTATGTGCATTTCTTCTGGCAT 1371
QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440


```

Db 1372 GGTGTTTCCAGGTCGTGTCACATATATGAAATAATCTGGACATTAAGTCGCCACAG 1431
Qy 441 LysSerLysGlnGlnAspSerThrTyProlLeuLysLeuValAlaSerGlyMet 460
Db 1432 AAGAGCAAGAACACAGGATTCTACTTACCTATTAAAGATGAGGTGGCAGTGGATG 1491
Qy 461 IleuValMetAlaPhePheLeuIleThrTyPheHisSerThrTyPheValSer 480
Db 1492 ATACTGTCAGGCTTTTTCATCACCACCTACGCTTTCACCTGAGTGGTACCAAGT 1551
Qy 481 GluAlaTyPheSerProSerLysValLeuSerAlaArgGlyGlyAspGlySerArg 500
Db 1552 GAAGCTTATCTTCTCCCTCCATTTACTGCTGCTGCTGGTGGGATGGCAGTAGGATC 1611
Qy 501 IlePheAspPheArgGluAlaTyPheTyPheLeuArgHisAsnThrProGluAspAla 520
Db 1612 ATTTTGTACTTCCGAGAGGCTATTATTTGCTCCGTCACAACTACCTCCAGAGATGCA 1671
Qy 521 LysValMetSerTrpTrpAspTyPheGlyTyPheGlnIleThrAlaMetAlaAsnArgThr 540
Db 1672 AAAGTCATGTCATGGTGGGATTTATGCTACCAAAATTAATGCAATGGCAAAATCGGACAA 1731
Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisLeuSerArgValGlyGluAlaMetAla 560
Db 1732 TTAGTGACAAATACACATGGAATATATACCAATATTTCTCGAGTAGGCGAGCAATGGCA 1791
Qy 561 SerThrGluGluLysAlaTyPheGluIleMetArgGluLeuAspValSerTyPheValLeuVal 580
Db 1792 TCCACAGAGAAAGGCTATGAATCATGAGGAGCTTGAATGCTCAGCTATGCTGCTGTC 1851
Qy 581 IlePheGlyGlyLeuThrGlyTyPheSerSerAspAspLeuAsnLysPheLeuTrpMetVal 600
Db 1852 ATTTTGGAGGCTTACTGGGTATTTCTTGGATGATATCAACAAATTTCTTGGATGCTC 1911
Qy 601 ArgIleGlySerThrAspThrGlyLysHisLeuLysGluAsnAspTyPheTyPhePro 620
Db 1912 CGGATTGGAGAGACACAGACAGAGAGGACACATTAAGAGAAATGACTACTATACTCTCT 1971
Qy 621 ThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeuMetTyP 640
Db 1972 ACTGGGAAATTCGTGTGATGCTGAGGTTCTCCGGTCTCTCACTGCTGCTATATGATC 2031
Qy 641 LysMetCysTyPheArgPheGlyGlnValTyPheGluAlaLysArgProProGlyPhe 660
Db 2032 AAAATGTGTTACTACCGCTTTGGGACAGGCTACACAGAGAGCCAAAGCGTCCACAGGCTTT 2091
Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
Db 2092 GACCGTGTTCGAATGCTGAGATTTGATTAAGACTTTGAGCTTGTGCTCTGGAGGAA 2151
Qy 681 GlyTyPheThrGluHisTrpLeuValArgIleTyPheLysValLysAspLeuAspAsnArg 700
Db 2152 GCGTATACCAAGAACACTGGCTAGTCAGGATATACAGGTAAAGAGCTCTGGATAATCGA 2211
Qy 701 GlyLeuSerArgThr 705
Db 2212 GCCTTGTCAAGGACA 2226

```

RESULT 8

```

US-10-417-375-99
; Sequence 99, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 5404

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-99
Alignment Scores:
Pred. No.: 0 Length: 5404
Score: 3597.00 Matches: 686
Percent Similarity: 99.13% Conservatives: 1
Best Local Similarity: 98.99% Mismatches: 6
Query Match: 97.30% Indels: 6
DB: 18 Gaps: 0
US-10-028-384-12 (1-705) x US-10-417-375-99 (1-5404)
Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyPheGlnLysGlnAspThrLeuLeuLysLeu 20
Db 136 ATGACTAAGTTTGGATTTTTCGATTTGCTCTATGAGAGAGGAGGACACACTTTTGAAGCTT 195
Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 196 CTCATTCTGTCAATGGCTGCTGTATTATCTCTCCACTGCTGTTGCTGTCTCTCTGAGA 255
Qy 41 PheGluSerValIleHisGluPheAspProTyPheAsnTyPheArgThrArgPheLeu 60
Db 256 TTTGAAAGTGTATCCAGTATGATCCGCTACTTTAATTATCGGACTACCAAGTTCCCTG 315
Qy 61 AlaGluGluGlyPheTyPheLysPheHisAsnTrpPheAspAspArgAlaTrpTyPheLeu 80
Db 316 GCTGAGGAGGGGTTTATAAATTCCTAATCTGTTGATGACCGAGCCTGTACCTTTG 375
Qy 81 GlyArgIleIleGlyGlyTyPheTyPheProGlyLeuMetIleThrSerAlaAlaIleTyP 100
Db 376 GACCAATCATTTGGAGGACAAATTTACCCAGGTTTATGATCACTCTGCTGCAATCTAC 435
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 436 CATGTACTCTCAATTTTTCACATCACCATCGACATTCGGAATGCTGTGTGTTCTCTGGCC 495
Qy 121 ProLeuPheSerSerPheThrSerIleValTyPheTyPheLeuLeuTyPheLysGluLeu 140
Db 496 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 555
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 160
Db 556 GCAGGGGCTGGGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 615
Qy 161 SerValAlaGlySerTyPheAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyP 180
Db 616 TCTGTGGCTGGCTCTCTATGATTAATGAAGGATTTGCCATCTTTTGCATGCTACTCACTAC 675
Qy 181 TyPheMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 676 TACATGTGATCAAGGACGAGTAAAGACTGGTTCCATCTGTGTTGGGAGCTAAGTGTGGCTT 735
Qy 201 AlaTyPheTyPheMetValSerSerTrpGlyGlyTyPheValPheLeuIleAsnLeuLeuPro 220
Db 736 GCTTATTTCTACATGCTCTCTGATGGGAGGTTATGTTCTCTGATCACTTAATTCCT 795
Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyPheValAlaTyP 240
Db 796 CTCACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 855
Qy 241 CysThrValTyPheLysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 856 TGTACTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 915
Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 916 CTTGCTCTCTTCTATCAGACACATGCGAGCTTTGGGCTCTTTGGTCTCTCTCTCTCTCT 975
Qy 281 AlaPheValAspTyPheLeuArgSerTyPheLeuAsnProGlnGlnPheGluValLeuPheArg 300
Db 976 GCCTTTGTGGATTACCTGCGCAGCAAGTTGAATCCACAAATTTGAAGTTCTTTTCCGG 1035

```


QY	301	SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr	320
DB	1036	AGCGTCATCTCTCGTAGCGTTGCTTCTCACCGTGGAGCTCTCTCATCTGACA	1095
QY	321	GlyIysIleSerProTTPThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys	340
DB	1096	GGAAAAATATCTCCCTGGACGGGGCGTCTTACTCGCTGGATCCCTCTTATGCTAAG	1155
QY	341	AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr	360
DB	1156	ACACATCTCCCATCATCTGCTCTGTGCTGACATCAGCCACAACTGCTCTCATAC	1215
QY	361	TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer	380
DB	1216	TATTTGACCTGGACGCTCTCGTCTTCATGTTCCAGTTCGCTCTATTACTGCTTAGC	1275
QY	381	AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer	400
DB	1276	AACCTGCTGATGCCCGGANTTTATCATCATGATGTTGTGACCAACGATGACTTTTCA	1335
QY	401	AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle	420
DB	1336	GCTGTATGGTGCCTCTAAATCTAGTGTGGCACCTGTTATGTGCAATCTCTCTGGCAT	1395
QY	421	GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys	440
DB	1396	GGAGTCTCCCAAGTCTCTCCACATACATGAGAACTCTGGACATAAGTCGTCCAGACAG	1455
QY	441	LysSerIysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet	460
DB	1456	AAGAGCAAGACCAACAGGATCCACCTACCTCTATTAAGAATGAGTGGCAAGTGGGATG	1515
QY	461	IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTripValThrSer	480
DB	1516	ATACTGGTCATGGCTTCTTCTCATCACTACACTCATCCCTTCATTCAACCTGGGTGACCA	1575
QY	481	GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle	500
DB	1576	GAGGCTACTCTTCTCGTCCATGTACTATCTGCCCGTGGGGATGGCAGTAGGATC	1635
QY	501	IlePheAspAspPheArgGluAlaTyrTyrTyrTyrLeuArgHisAsnThrProGluAspAla	520
DB	1636	ATATTGTGATGACTCCGAGGAAGCATATATTGGCTTCGTCTAATAATACCCAGAGGATCG	1695
QY	521	LysValMetSerTTPTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle	540
DB	1696	AAAGTCAATGCTCTGGTGGGANTATGGCTATCAGATTACAGCTATGGCAACCGAACCA	1755
QY	541	LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla	560
DB	1756	TTAGTGGACAATAACATACGGAATAATACCCATATTTCTCGATAGGCGAGCAATGGCG	1815
QY	561	SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal	580
DB	1816	TCCACAGAGGAAAAGGCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTGC	1875
QY	581	IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal	600
DB	1876	ATTTTGGAGCCCTACTGGGTATTCCTGTGATGATATCAACAAGTTTCTTTGGATGGCT	1935
QY	601	ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro	620
DB	1936	CGAATTGGAGGAGCAGACATACAGCAACAACATATCAAGAGGAATGACTATTATCTCCA	1995
QY	621	ThrGlyGluPheArgValAspArgGlySerProValLeuLeuAsnCysLeuMetTyr	640
DB	1996	ACTGGGAGTTCCTGGTGGACCGTGAAGGTCTCCAGTCTGCTCACTGCTCATGTAC	2055
QY	641	LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe	660
DB	2056	AAAGATGCTTACTATCCCTTTCAGAGTTTACAGAAAGCAACGCTCCCTCAAGGTTT	2115

QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680

Db 2116 GACCGTGTCGCAATGCTGAGATGGGAATNAAGACTTTGAGCTTGATCTCTGGAGAA 2175

QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLys 693

Db 2176 GCATATACCAGAACATTGGCTGGTCAGSATATACAAG 2214

RESULT 9

US-10-417-375-97

; Sequence 97, Application US/10417375

; Publication No. US20040219528A1

; GENERAL INFORMATION:

; APPLICANT: David W. Morris

; APPLICANT: Marc Malandro

; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer

; FILE REFERENCE: 529452001600

; CURRENT APPLICATION NUMBER: US/10/417,375

; CURRENT FILING DATE: 2003-04-15

; NUMBER OF SEQ ID NOS: 176

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 97

; LENGTH: 5827

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-417-375-97

Alignment Scores:

Pred. No.: 0 Length: 5827

Score: 3597.00 Matches: 686

Percent Similarity: 99.13% Conservative: 1

Best Local Similarity: 98.99% Mismatches: 6

Query Match: 97.30% Indels: 0

DB: 18 Gaps: 0

US-10-028-384-12 (1-705) x US-10-417-375-97 (1-5827)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20

Db 136 ARGACAAAGTTTGGAATTTTGGCGAATGTCTCTATGAGACGAGGACACACTTTTGAAGCTT 195

QY 21 LeuIleLeuSerMetAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40

Db 196 CTCATTCTGTCAATGGCTGCTGTATTATCTCTCCACTGCTGTTTGTCTGCTCCTGAGA 255

QY 41 PheGlnSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60

Db 256 TTGAAAGTGTAATCCAGTAGTTTATCCGTACTTTAATATCGGACTACCAAGTTTCTG 315

QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspAlaIleTyrProLeu 80

Db 316 GCTGAGGAGGGGTTTATATAAATCCATAACTGGTTTGATGACCGAGCGCTGTACCCCTTG 375

QY 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100

Db 376 GGACGAATCATTTGGAGAACAAATTTACCAAGGTTTAATGATCACCTCTGCTGCATCTAC 435

QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120

Db 436 CATGTACTCCATTTTTCACATCCACCATCGCAATTCGGAATGCTGTGTGTTCCTCGCC 495

QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140

Db 496 CCTCTCTCTCTCTCTCCATCCACCACCATGCTGCTACCACTTACCAAGAGSCTCAAGGAT 555

QY 141 AlaglyAlaglyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160

Db 556 GCAGGGGGCTGGCTTCTGTGCTGCCATGATGTGTGTAGTTCTCTGGATATATCTCCCGA 615

QY 161 SerValAlaglySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180

Db 616 TCTGTGGCTGGCTCTATGATATATGAAGGATTCGATCTTTTGCATGCTATCACCTAC 675

QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
 DB 676 TACATGTGGATCAAGGCAGTAAGAGCTGGTTCCATCTGTGGGCAGCTAAGTGTGCCCTT 735
 QY 201 AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuLeuPro 220
 DB 736 GCTTATTCTTACATGGTCTCGTATGGGAGGTTATGTTCTCTGATCAACTTAATTCCT 795
 QY 221 LeuHisValLeuValLeuMetLeuThrGlyATGpPheSerHisArgIleTyrValAlaTyr 240
 DB 796 CTCACGTCCTCGTGTGATGCTCACAGGCGCTTCTCTCACCGGATCTATGGCCCTAC 855
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 DB 856 TGTACTGTTTACTCCCTGGGCACTATATCTTCTATGAGATCTCTTGTGGGTTTCCAG 915
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 DB 916 CCTGTCCTTTCATCAGAGCACATGGCAGCCCTTTGGGGTCTTGGTCTCTGCCAGATCCAT 975
 QY 281 AlaPheValAspTyrIleuArgSerLysLeuAspProGlnGlnPheGluValLeuPheArg 300
 DB 976 GCTTTTGTGGATACCTCCGACGACAGTGTGAATCCACAAATTTGAAGTTCTTTTCCGG 1035
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 DB 1036 AGCGTCATCTCTGGTAGGCTTTGTCCTTCTCACCGTGGAGCTCTCTCATGTGTGACA 1095
 QY 321 GlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 DB 1096 GGAATAATATCTCCCTGGAGGGGGTTCCTCTCTGCTGTGATCCCTCTTATGCTAAG 1155
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrTrpSerSerTyr 360
 DB 1156 AACACATCCCATCATTTGCTGTCTGTGACATCAGCCACCAACCTGGTCCCTCATAC 1215
 QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 DB 1216 TATTTGACCTCGACGCTCTCTGCTCTGCTCTGCTCTGCTCTCTCTCTCTCTCTCT 1275
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1276 AACCTGTCTGATGCGCGGATTTTATCATCATGTATGGTGTGACCAAGCATGTACTTTCA 1335
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1336 GCTGTATGCTGCTTATGCTAGTGTGGACCTGTTATGTGCACTCTCTCTGCGATT 1395
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 DB 1396 GGAGTCTCCAGGTCGTGTCCATCATCATGAAGATCTGGACATAGTGTCCAGACAAAG 1455
 QY 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 DB 1456 AAGAGCAGAGCAGCAGAGATCCACCTACCTATTAAGAAAGAGTGGCAAGTGGGATG 1515
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThrSer 480
 DB 1516 ATACTGTGTCATGGCTTTCTTCTCATCACCTACACCTTTCTTCAACCTGGGTGACCAAGT 1575
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyLysAspGlySerArgIle 500
 DB 1576 GAGGCTACTCTTCTCGTCCATGTTACTATCTGCGGTGGTGGGATGCGATGAGATC 1635
 QY 501 IlePheAspAspPheArgGluAlaTyrTrpLeuArgHisAsnThrProGluAspAla 520
 DB 1636 ATATTGTATGATCTCCAGAGCAGATATTATGCTCTTCTGATATATCTCCAGAGATGCG 1695
 QY 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 DB 1696 AAGGTCATGCTCTGTGGGATATGCTATACAGTATACAGTATGCAACCCGACAAATT 1755
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560

DB 1756 TTAGTGGCAATAACACATGGAATAATACCATATTTCTCGAGTAGGCGAGCAATGGCG 1815
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1816 TCCACAGAGGAAAGCCATGATGATCATGAGGAGCTCGATGTGATGTGCTGGTC 1875
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 DB 1876 ATTTTGGAGGCTCCTGCGGTATCTCTGATGATATCAACAGTTCTTTTGGATGGTC 1935
 QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrTrpPro 620
 DB 1936 CGGATTTGGAGGAGCACAGATACAGGCAACATATCAAGGAGATCTATTATATCTCCA 1995
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 DB 1996 ACTGGGAGTTCCTGTCGACCGTGAAGTTCTCCAGTGTCTCACTCCCTCATGTATC 2055
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 DB 2056 AGATGTGTACTATCTGCTTTGGACAGTTTACACAGAGCCAGCGTCTCTCCAGCTTT 2115
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 DB 2116 GACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGCTCGGAGAA 2175
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLys 693
 DB 2176 GCATATACACAGAACATTTGCTGTCAGGATATACAAG 2214

RESULT 10

US-10-425-115-150745
 ; Sequence 150745, Application US/10425115
 ; Publication No. US20040214272A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants
 ; FILE REFERENCE: 38-21(53222)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369326
 ; SEQ ID NO 150745
 ; LENGTH: 2839
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: MRR4577_69000C.1
 US-10-425-115-150745

Alignment Scores:
 Pred. No.: 1,178-222 Length: 2839
 Score: 2258.00 Matches: 420
 Percent Similarity: 75.91% Conservative: 122
 Best Local Similarity: 58.82% Mismatches: 142
 Query Match: 61.08% Indels: 30
 DB: 18 Gaps: 7
 US-10-028-384-12 (1-705) x US-10-425-115-150745 (1-2839)

QY 7 LeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeuLeuLeuLeuSerMetAla 26
 DB 254 CTGCGCTCAAGCAAGCAAGCAGGAGCTCTGCTCCGCTCGGCTGGCGCTCATC 313
 QY 27 AlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValIleHis 46
 DB 314 TACGTGCTGCTCTGCGCTCGCTCTCTCTCCGTCTCCGTACGAGTCCATGATCCAC 373
 QY 47 GluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGluGlyPheTyr 66

Db 374 GAGTTCGACCCCTTACTCAACTACCGCACCGCTCTCTCTCAACGAGACCGCTTCAGC 433
Qy LysPheHisAsnTrpPheAspArgAlaTrpTyrProLeuGlyArgIleGlyGly 86
Db 434 GAGTTCGACCCCTTACTCAACTACCGCACCGCTCTCTCTCAACGAGACCGCTTCAGC 493
Qy ThrIleTyrProGlyLeuMetIlePheSerAlaAlaIleTyrHisValLeuHisPhe 106
Db 494 AGCTCTACCCGGGCTCATGTGACCGCCGGCTGTCTCACCGCCCTCTCCCGCGCTC 553
Qy HisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerPhe 126
Db 554 TACCTCACCGCTCCACATCGGTGAGGTCTCGCTCCCTCACCGCGCTCTCTCTCGCTGCAAC 613
Qy ThrSerIleValThrLeuLeuThrLysGluLeuLysAspAlaGlyIleLeu 146
Db 614 ACCAGCTGCTGCTACGCTTCGCGCCGAGATGCGGACTCCGGGGCGGCTGCTC 673
Qy 147 AlaAlaMetIleAlaValProGlyTyrIleSerArgSerValAlaGlySerTyr 166
Db 674 GCGCTGCGCTCATCGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 733
Qy 167 AspAsnGluGlyIleAlaPheCysMetLeuLeuThrTyrTyrMetTyrIleLysAla 196
Db 734 GACACAGGGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 793
Qy 187 ValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeuAlaTyrPheTyrMetVal 206
Db 794 GTTAACACTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 853
Qy 207 SerSerTrpGlyTyrValPheLeuIleAsnLeuIleProLeuHisValLeuValLeu 226
Db 854 TCGGATGGGAGGCTAGTGTTCATCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 913
Qy 227 MetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrValTyrCysLeu 246
Db 914 CTGTGCACTGGGAGTACTCGCAGAGGCTACTACGCTGCTGCTGCTGCTGCTGCTGCTGCT 973
Qy 247 GlyThrIleLeuSerArgGlnIleSerPheValGlyPheGlnProValLeuSerSerGlu 266
Db 974 GGATGTGCTTGCATCGAGATTCGTTTGTGATTCAGATTCAGATTCAGATTCAGATTCAG 1033
Qy 267 HisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeu 286
Db 1034 CACATGGCTGCTATGGAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1093
Qy 287 ArgSerLysLeuAsnProGlnPheGluValLeuPheArgSerValIleSerLeuVal 306
Db 1094 AAATATCTGCTAAATGATGCCAGA-----CTATTCAGTCAATCTCTGCAATTACC 1144
Qy 307 GlyPheValLeuThrValGlyAlaLeuLeuMet-----LeuThrGlyLys 322
Db 1145 CTCACATGCTGATTACTATGGCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1204
Qy 323 IleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLysAsnAsn 342
Db 1205 ATCTCCCTTGGACAGGACGGTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1264
Qy 343 IleProIleAlaSerValSerGluHisGlnProThrThrThrThrThrThrThrThrThrThr 362
Db 1265 ATACCAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1324
Qy 363 AspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSerAsnLeu 382
Db 1325 GACTTCCACATCT 1384
Qy 383 SerAspAlaArgIlePheIleLeuMetTyrGlyValThrSerMetTyrPheSerAlaVal 402
Db 1385 TCAGATGCCACAAATATTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1444
Qy 403 MetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIleGlyVal 422

1445 ATGTCGCGTTAATCTTGTGGCAGCACCGCGTGTTCCTTATAGTCTATTGCTGCA 1504
Qy 423 SerGlnValLeuSerThrTyrMetLysAsnLeu---AspIleSerArgProAspLysLys 441
Db 1505 TCTGCTCAATA-----AAAAACCTAACCACTTTGATCAGGACAAAGAGCAGA 1552
Qy 442 Ser-----LysLysGlnGln 446
Db 1553 AGTCCACAGACTACTTCTGGGAAAAACACTGGCTCAAGTCAGCTCCAAAGGGTGCAGTT 1612
Qy 447 AspSerThrTyrProIleLysIleGluValAlaSerGlyMetIleLeuValMetAlaPhe 466
Db 1613 GATCAATCTTTACCATTTCCACCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1672
Qy 467 PheLeuIleThrTyrThrPheHisSerThrTyrValThrSerGluAlaTyrSerSerPro 486
Db 1673 TTGCTCAGTAGTATGCCATACACTGCACCTTGGGTGCATCTGAGGCTTACTCTTCTTCCA 1732
Qy 487 SerIleValLeuSerAlaArgGlyAspGlySerArgIleIlePheAspPheArg 506
Db 1733 TCTATAGTTTGGCGCAGGGGTCTAATATGGTGAAGAGTCTATCTTGTGATTTATCGT 1792
Qy 507 GluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrp 526
Db 1793 CAGGCACTACTATGCTTCTGTCAGACACGCTCTCTGATGCTAAGATTTATGCTGCTGG 1852
Qy 527 AspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThr 546
Db 1853 GACTATGATACCAATATACAGTATGGTAAACAGAACTGTATTGTTGATATACATACA 1912
Qy 547 TrpAsnAsnThrHisIleSerArgValGlyGluAlaMetAlaSerThrGluGluLysAla 566
Db 1913 TGGATTAACACACATAGTACTGATGAGGAGCTATGCTATCTTATGAGATGAGCA 1972
Qy 567 TyrGluLeuMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThr 586
Db 1973 TATGAGATCATGCACTGCTGATGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2032
Qy 587 GlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThr 606
Db 2033 GGATATCTTCAGATGATATCAACAGTCTTATGATGATGATGATGATGATGATGATGATGATG 2089
Qy 607 AspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThrGlyGluPheArgVal 626
Db 2090 ---GTTTTCTCTGTAATCAAGAGCCAGTACTCTGTT---AATGGGGAGTATCTGTT 2143
Qy 627 AspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArg 646
Db 2144 CACAAGGGGGCAGCACCCCAAAATGCTGAATGCTGAATGCTGAATGCTGAATGCTGAATGCTGA 2203
Qy 647 PheGlyGlnValTyrThrGluAlaLysArgProProGlyPheAspArgValArgAsnAla 666
Db 2204 TTTGAGAACTTACACAGAAATATGAAAAACCTCCAGGGTATGATGCTGTACGAGGAGTG 2263
Qy 667 GluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyrThrThrThrGluHis 686
Db 2264 GAGATTGGCAACAAGACATAAGCTTGAATCTTGGAGGAGGATTTACACTTCAACTTCAAC 2323
Qy 687 TrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
Db 2324 TGGATAGTGGCATATCAAGGTGAACCCCAAGAACACAGG 2365

RESULT 11
US-10-425-114-14408
; Sequence 14408, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaka, Jack E
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 14408
; LENGTH: 2681
; TYPE: DNA
; ORGANISM: Zea mays

; FEATU: OTHER INFORMATION: Clone ID: LIB189-017-A10_FLI
US-10-425-114-14408

Alignment Scores:

Pred. No.: 1,94e-222 Length: 2681
Score: 2255.50 Matches: 420
Percent Similarity: 75.88% Conservative: 121
Best Local Similarity: 58.91% Mismatches: 143
Query Match: 61.01% Indels: 29
DB: 16 Gaps: 7

US-10-028-384-12 (1-705) x US-10-425-114-14408 (1-2681)

```

QY 7 LeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeuLeuLeuSerMetAla 26
DB 221 CTGGCGCTCAAGACGAAAGCAGCAGGAGCTCTGCTCCGCTCTCGCGCTCGCGCTATC 280
QY 27 AlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValIleHis 46
DB 281 TACGTGCTGCCCTTCGGCGCTCGCGCTCTCTCGCTCTCGCTACGAGTCCATGATCCAC 340
QY 47 GluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGlyPheTyr 66
DB 341 GAGTTCGACCGCTTACCTTCAACTACCGCACCACCGCTCTCTCTCACCAGAACCGCTTCAGC 400
QY 67 LysPheHisAsnTyrPheAspAspArgAlaTyrTyrProLeuGluValIleGlyGly 86
DB 401 GAGTTCGAACTGGTTCGACTCGAGAGCTGGTACCGCTCGCGCGCTCTCGCGCGG 460
QY 87 ThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyrHisValIleHisPhe 106
DB 461 ACGTCTACCGCGCTCGATGTTGATGGTGAACCGCGCGCTCTCTCAGCGCTCTCTCGCGCGCTC 520
QY 107 HisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerPhe 126
DB 521 TACCTACCGTCCACATCCGTGAGGCTCGGTCTCTCACCAGCGCTCTCTCTCGCTCCAC 580
QY 127 ThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAspAlaGlyAlaGlyLeuLeu 146
DB 581 ACCACGCTCGCTCGCTACGCTTCGGCGCGAGATATGGAGCTCCGCGCGCGCGCTCGCTC 640
QY 147 AlaAlaAlaMetIleAlaValValProGlyTyrIleSerArgSerValAlaGlySerTyr 166
DB 641 GCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 700
QY 167 AspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetIlePheLysAla 186
DB 701 GACAAACAGGCGCTCGCGCTCTCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 760
QY 187 ValLysThrGlySerIleCysTyrAlaAlaLysCysAlaLeuAlaTyrPheTyrMetVal 206
DB 761 GTTAACACTGGTCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 820
QY 207 SerSerTyrGlyTyrValPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 226
DB 821 TCGCATGGGAGGCTACGTTGTTTCATCATCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 880
QY 227 MetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrValTyrCysLeu 246
DB 881 CTTCCTACTGGAGGTACTCGCAGAGGCTCTACGTGCGCTCACTCAATGATGATGCTT 940
QY 247 GlyThrIleLeuSerArgGlnIleSerPheValGlyPheGlnProValLeuSerSerGlu 266

```

```

DB 941 GGATGTTGCTTGCATGAGATTCGTTTTCGATTCAGCATGCGATCGCGAGAG 1000
QY 267 HisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeu 286
DB 1001 CACATGGCTGCTATGGAGTCTTCTCTCTGTCAGGTTTCTCTCTCTCTCTCTCTCTCTCT 1060
QY 287 ArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuVal 306
DB 1061 AAATATCTGCTAAATGATGCCAGA-----CTATTCAAGTCATTCTTCGAATTACC 1111
QY 307 GlyPheValLeuLeuThrValGlyAlaLeuLeuMet-----LeuThrGlyLys 322
DB 1112 CTCACATGCTGATTACTATTGGCACCTGCTCTTGACCTGCTACTGATCTGCTTAC 1171
QY 323 IleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLysAsnAsn 342
DB 1172 ATCTCCCTTGGACAGGACGCTTTTACTCCTCTGCTGATCTCTCTCTCTCTCTCTCTCTCT 1231
QY 343 IleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyrTyrPhe 362
DB 1232 ATACCAATCATCGCATCTGTTCTGAGCATCAGCCACAGCATGCTCTCTCTCTCTCTCT 1291
QY 363 AspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSerAsnLeu 382
DB 1292 GACTTCCACATCTCTTCTTGTCTCCAGCTGGCTTTATTTCTGTTTCAAGCGCTG 1351
QY 383 SerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSerAlaVal 402
DB 1352 TCAGATGCCACGATATTATTGTTATGATGATGCTCAAGATGATGTTCTTCTGCTG 1411
QY 403 MetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIleGlyVal 422
DB 1412 ATGGTCCGGTTAAATTCCTTGTGGCAGCACCAGCGCTTATGCTTATGCTGATGCTGCA 1471
QY 423 SerGlnValLeuSerThrTyrMetLysAsnLeu---AspIleSerArgProAspLysLys 441
DB 1472 TCTGCTACAATA-----AAAATCTAACCACTTTTATGCTCCGACAAAGAGCAAA 1519
QY 442 Ser-----LysLysGlnGlnAsp 447
DB 1520 AGTCCACAGACTACCTGGAAGAAACAACTGGTTCAAAGCAGGTGCAAGGAGGAGCAGTTGAT 1579
QY 448 SerThrTyrProIleIleGluValAlaSerGlyMetIleLeuValMetAlaPhePhe 467
DB 1580 CAATCTTTACCATCTCCACACACATGCTGCTATGCTCTCTCTCTCTCTCTCTCTCTCT 1639
QY 468 LeuIleThrTyrThrPheHisSerThrTyrValThrSerGluAlaTyrSerSerProSer 487
DB 1640 CTCAGTAGTATGCAATACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1699
QY 488 IleValLeuSerAlaArgGlyLysAspGlySerArgIleIlePheAspPheArgGlu 507
DB 1700 ATAGTTTGGCTGCAAGGGGTCATAATGCTGGAAGAGTCATCTTTGATGATTATCGTGAG 1759
QY 508 AlaTyrTyrTyrLeuArgHisAsnThrProLysAspAlaLysValMetSerTyrTrpAsp 527
DB 1760 GATACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1819
QY 528 TyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrp 547
DB 1820 TATGGATACCAATATACAGCTATGGGTAAACAGAACTGTTATTGTTGATAACATATACATG 1879
QY 548 AsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyr 567
DB 1880 AATAACACACACATAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 1939
QY 568 GluIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGly 587
DB 1940 GAGATCATCGACTCTGATGTAACCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1999
QY 588 TyrSerSerAspAspIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThrAsp 607

```

Db 2000 TATTCCTCAGATGATATCAACAGTCTTATGATGGTACGTATGCTGGAGGA----- 2053
QY 608 ThrGlyLysHslelleLysGluAsnAspTyrThrProThrGlyGluPheArgValAsp 627
Db 2054 GTTTTCTCTGTAATCAAGAGCCAGATACCTTGT---AATGGGAGTATCGTATTGAC 2110
QY 628 ArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPhe 647
Db 2111 AAGGGGGGAGCCCAAAATGCTGACTGCTTAATGTAACAGCTTTGTTACTACCGATT 2170
QY 648 GlyGlnValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaGlu 667
Db 2171 CGAAGAACTTACCACAGATATGGAACCTCCAGGGTATGATCGTACGAGAGTGGAG 2230
QY 668 IleGlyAsnLysAspPheGluLeuValLeuGluGlyTyrThrThrGluHisTyr 687
Db 2231 ATGGCAACAGACATTAAGCTTGATACCTTGGAGGAGGCAATTACACTTCAACTGG 2290
QY 688 LeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
Db 2291 ATAGTGGCATATACAGGTGAACCCCAAGAACAGG 2329

RESULT 12

US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)
; OTHER INFORMATION:

US-10-028-384-1

Alignment Scores:
Pred. No.: 1,3e-221 Length: 2481
Score: 2247.00 Matches: 427
Percent Similarity: 73.99% Conservative: 102
Best Local Similarity: 59.72% Mismatches: 150
Query Match: 60.78% Indels: 36
DB: 15 Gaps: 7

US-10-028-384-12 (1-705) x US-10-028-384-1 (1-2481)

QY 16 ThrLeuLeuLysLeuLeuLeuSerMetAlaValLeuSerPheSerThrArgLeu 35
Db 208 TCGCTTCCTCCTCACCATCTCTTCTGGCCCTGGCTTGGCCGCTTCAGCTCGCGCCTC 267
QY 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
Db 268 TTCGCGCTCATCGCTTCGAAGCATCATCCAGGTTCCAGCCGCTGTTAACTATAGA 327
QY 56 ThrThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsnTyrPheAspArg 75
Db 328 TCACACATCATCTTGCATCTCATGGGTCTATGAATTTTAAATTTGGTTTGATGAAGA 387
QY 76 AlaTyrProLeuGlyArgIlelleGlyThrIleTyrProGlyLeuMetIleThr 95
Db 388 GCATGTAATCCACTAGGAAGATAGTGGTACTGTTTACCAGGTTGATGATAACC 447
QY 96 SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal 115

Db 448 GCTGGCCTTATTCATTGGATTTTAAATACATTAACATGTTCCACATAGAGACGTA 507
QY 116 CysValPheLeuAlaProLeuPheSerPheThrSerIleValThrTyrLeuLeuThr 135
Db 508 TGTGTGTTCTTGTCACCAACTTTTAGCGGCTTATACATCTATCTACTTCTCTGCTTACA 567
QY 136 LysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValPro 155
Db 568 AGAAGACTTTTGAACCAAGGAGGAGGACTTTTAGCTGCTGTTTATTCTATTGTACCA 627
QY 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
Db 628 GGCTACATATCTCGCTCAGTAGCTGCATCTTTTATATAATCAAGGCATTGCTATTTTGA 687
QY 176 MetLeuLeuThrTyrTyrMetTyrIleLysAlaValLysThrClySerIleCysTrpAla 195
Db 688 CTTCACTTCATACATCTATTATGGTAAATCTGTAAACCTGGTCACTTTTGGACA 747
QY 196 AlalysCysAlaLeuAlaTyrPheTyrMetValSerSerTyrPolyGlyTyrValPheLeu 215
Db 748 ATGTGCTGCTCTATCTCTATATGCTCTGCTTGGGCTGCTTATGTTATTTATC 807
QY 216 IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
Db 808 ATCAATCTTATCTCCATGCAATGTTTGTGTTGTTACTGATGAGAGATACAGCAAAAGA 867
QY 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
Db 868 GTCTACATAGCATATAGCATCTTCTACATTTGCTGGTTTAAATATTAATCAATGAGATACCT 927
QY 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaClyPheGlyValPheGly 275
Db 928 TTTGTGGATTCCAGCCATCAAGAACAGTGAACATGACGCTCCAGCTGCTGCTTTGCA 987
QY 276 LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPhe 295
Db 988 TTGCTGCAAGCTTATGCTTCTTCTGAGATCTGAGAGCCGATTAACAAACAAAGAGTTC 1047
QY 296 GluValLeuPheArgSerValIleSerLeu---ValGlyPheValLeuLeuThrValGly 314
Db 1048 CAGACCTTTTCTTTTGGGTGATCTACTAGCTGCGAGTGTGTGTTCTTAGTGTC--- 1104
QY 315 AlaLeuLeuMetLeuThrGlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeu 334
Db 1105 ATCTATTGACTTATACAGGTTTACATTTGCACCATGAGTGGCAGGTTTATTTCATTGCG 1164
QY 335 AspProSerTyrAlaLysAsnAsnIleProIleAlaSerValSerGluHisGlnPro 354
Db 1165 GATCTGGGTATGCAAAAATACATTCCTCAATTTATTCATCAGTGTCTGAGCATCACT 1224
QY 355 ThrThrPheSerTyrTyrPheAspLeuGlnLeuValPheMetPheProValGly 374
Db 1225 ACGACTTGGGTGCTCTTCTTCTTGTATCTACATATCTTGTATGATACCTTCCAGCAGGC 1284
QY 375 LeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyVal 394
Db 1285 CTTTGGTCTGCATCAAAAATATCAAGCATGAAGAGTATTTGCTCTTATGCAATC 1344
QY 395 ThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMet 414
Db 1345 AGTGTGCTACTTGTGCTGAGTGCATGCTGCGACTGATGTTGACTTTGACTTCCAGTCGTG 1404
QY 415 SerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAsp 434
Db 1405 TGTATGCTGCTGCAATGCTTTTCAATGTTTGTGAGCACTATTGTTGGGAT---GAC 1461
QY 435 IleSerArgPro----- 438
Db 1462 ATGAAAAGGGAATAATCCACTGTGGAGCAGCAGTATGAGGATGACAAAAGAACCAA 1521
QY 439 -----AspLysLysSerLysLysGlnAspSerThrTyrProIleLysIle 454

```

Db 1522 GCAATTTGTATGATAGGAGGTAAGTACAGAAACATGCACTGACAGGAAAAAACT 1581
Qy 455 Glu-----ValAlaSerGlyMetLeuValMetAlaPhePheLeu 468
Db 1582 GAAGAGGATTAGGCCCTTAATAAAGCAATGTCACCATGTTGATGCTGATGCTATTG 1641
Qy 469 ileThrThrPheHisSerThrTrpValThrSerGluAlaThrSerSerProSerile 488
Db 1642 ATGATGTTTGTCTCCACTACTGCTGGTCACAGCAATGCTTCTTCTAGTCCAGAGTGA 1701
Qy 489 ValLeuSerAlaArgGlyGlyAspGlySerArgGlyleilePheAspAspPheArgGluAla 508
Db 1702 GTCTGGCCCTCATACATCATGATGTCAGCAGCAATATCTTAGATGANTTTAGAGAAGCT 1761
Qy 509 TyrThrTrpLeuArgHisAsnThrProGluAspAlaValMetSerTrpTrpAspTyr 528
Db 1762 TACTTTTGGCTAGGCAAAATACAGATGAACATGACGAGTAAATGTTCTGTGGGATTA 1821
Qy 529 GlyTyrGlnleThrAlaMetAlaAsnArgThrileLeuValAspAsnThrTrpAsn 548
Db 1822 GCCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGCTGGTGAATAAACACCTGGAAT 1881
Qy 549 AsnThrHisleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGlu 568
Db 1982 AACAGCCACATAGCACTGGTGGAAAGCTATGCTCTTAATGAACAGCAGCCTATAA 1941
Qy 569 ileMetArgGluLeuLeuAspValSerTyrValLeuValilePheGlyGlyLeuThrGlyTyr 588
Db 1942 ATCATGAGGACTCTAGATGATGATATGTTTGGTATTTTGGAGGGGTTATTGGCTAT 2001
Qy 589 SerSerAspAspileAsnLysPheLeuTrpMetValArgileGlyGlySerThrAspThr 608
Db 2002 TCTGCTGATGATATCAACAATTTCTCTGATGGTTAGATAGCT-----GAA 2049
Qy 609 GlyLysHis-----ileLysGluAsnAspTyrTyrThrProThrGlyGluPheArg 625
Db 2050 GGAGAACATCCCAAGACATTCGGAAAGTGAATATTTTACCCCAAGGAGAAATTCGCT 2109
Qy 626 ValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCystTyrTyr 645
Db 2110 GTACACAAAGCAGGATCCCTACTTTGTTGAATTCCTTATGATATAAATGTCACTAC 2169
Qy 646 ArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPheAspArgValArgAsn 665
Db 2170 AGATTGGAGAAATGCGAGCTGGATTTTCGTACACCCCGGAGTTTTCGACGACAGTAA 2229
Qy 666 AlaGluileGlyAsnLysAspPheGluLeuAspValLeuGluGluGlyTyrThrThrGlu 685
Db 2230 GCTGAGATTGGAATAGGACATTAATCAACATTTGGAAGAGCCCTTACATACAGA 2289
Qy 686 HistTrpLeuValArgileTyrLysValLysAspLeuAspAsnArg 700
Db 2290 CACTGGCTTTGATAGGATATATAAGTAAAGCAGCCTGATAACAGG 2334

```

RESULT 13

US-10-028-384-3

Sequence 3, Application US/10028384

Publication No. US20030148285A1

GENERAL INFORMATION:

APPLICANT: COMPATIGENE INC.

APPLICANT: PERREAU, Claude

APPLICANT: MCBRIDE, Kevin

FILE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy

FILE REFERENCE: 5600-74

CURRENT APPLICATION NUMBER: US/10/028,384

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 13

SEQ ID NO 3

LENGTH: 2710

TYPE: DNA

ORGANISM: Mus musculus

PUBLICATION INFORMATION:

```

; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)..(2469)
US-10-028-384-3

Alignment Scores:
Pred. No.: 1,628-220 Length: 2710
Score: 2237.00 Matches: 424
Percent Similarity: 73.57% Conservative: 102
Best Local Similarity: 59.30% Mismatches: 153
Query Match: 60.51% Indels: 36
DB: 15 Gaps: 7

US-10-028-384-12 (1-705) x US-10-028-384-3 (1-2710)

Qy 16 ThrLeuLeuLysLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu 35
Db 275 TGTGTGCTCTCTTCCATCTCTTCTGCGCTGCTGCGCGCTTCAGCTCGCGCTC 334
Qy 36 PheAlaValLeuArgPheGluSerValileHisGluPheAspProTyrPheAsnTyrArg 55
Db 335 TTGCGCGCTCATCCGCTTCGAGGACATCATCCAGGATTCGACCGCTGCTTAACATAGA 394
Qy 56 ThrThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsnTrpPheAspAspArg 75
Db 395 TCACACATCATCTTGCATCTCATGATCTCATGATTTCTAATGTTGTTGTAAGA 454
Qy 76 AlaTyrProLeuGlyArgileGlyGlyThrileTyrProGlyLeuMetileThr 95
Db 455 GCATGTGATCCCACTGGGAGAAATAGTGGTGGCACCCTTTACCAGGTTGATGATAACA 514
Qy 96 SerAlaAlaileTyrHisValleuHisPhePheHisleThrileAspilleArgAsnVal 115
Db 515 GTGCGCTTATTCATGATGATTTTAAACATATGACATTAACAGTTCACATTAAGAGATG 574
Qy 116 CysValPheLeuAlaProLeuPheSerSerPheThrSerileValThrTyrLeuLeuThr 135
Db 575 TGTGTATTCTTGTGACCAACTTTTAGCGGCTTACATCCATATCTAGTTCCTGCTAACT 634
Qy 136 LysGluLeuLysAspAlaGlyValaGlyLeuLeuAlaAlaMetileAlaValValPro 155
Db 635 AGAGAACTGTGACCAAGGAGGAGGACTTCTAGCTGCTGCTTCATGCTATGCTGACCA 694
Qy 156 GlyTyrileSerArgSerValAlaGlySerTyrAspAsnGluGlyleAlailePheCys 175
Db 695 GGTACATATCTCGGTGAGTGGCGGATCTTGTATTAATGAAGGATTCGCAATTTTTCG 754
Qy 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaVallyThrGlySerileCystTrpAla 195
Db 755 CTTCAGTTCATCTACTTATGAGTAAAGTCTGTGAAGCCGCTGTGTTCTGTCGACA 814
Qy 196 AlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeu 215
Db 815 ATGTGCTGCTGCTTGTCTATATTTTACATGCTCTCTGCTGGGAGGATTTATGTTTCATC 874
Qy 216 ileAsnLeuileProLeuHisValleuValleuMetLeuThrGlyArgPheSerHisArg 235
Db 875 ATCAACCTCATCCCTCTCCATGTTGTGTTGTGTTGTGTCGATGACGAGGTACAGACAGA 934
Qy 236 ileTyrValAlaTyrCysThrValTyrCysLeuGlyThrileLeuSerArgGlnileSer 255
Db 935 GTCTACATAGCATATAGCACTTTGTACATTTGTGGGTTTAAATATTATTCATGACAGATAC 994
Qy 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
Db 995 TTTTGGGATTTACGCAATCAGAACAGGACATGCGAGCTGCGAGGTGCTTTGCG 1054
Qy 276 LeuCysGlnileHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPhe 295
Db 1055 CTGCTGCAAGCTTACGCTTTTTCAGTATCTGAGAGCCGCTTGACAAACAGGAGTTC 1114
Qy 296 GluValLeuPheArgSerValileSerLeu--ValGlyPheValleuLeuThrValGly 314

```

Db 1115 CAGACCCCTTTCTTTTGGGTGCTCTACTAGCTGACGGCGTGTGTTCTTCTAGTGTCT 1171
Qy 315 AlaLeuLeuMetLeuThrGlyValSerProThrThrThrGlyValArgPheThrSerLeu 334
Db 1172 ATCTATCTGACATACACAGGTATATTCACCATGGAGTGGCGAGTTTATTCACATGG 1231
Qy 335 AppProSerThrAlaLysAsnAlaSerProThrThrThrGlyValSerGluHisGlnPro 354
Db 1232 GATACCTGGGTATGCAAAATACACATTCATTCATTCATTCATTCATTCATTCATTCAT 1291
Qy 355 ThrThrThrSerSerThrPheAspLeuGlnLeuLeuValPheMetPheProValGly 374
Db 1292 AGCATATGGGTGCTTTCTTTCTTTGATCTACATTCATTCATTCATTCATTCATTCAT 1351
Qy 375 LeuThrThrCysPheSerAsnLeuSerAspAlaAlaArgPheThrThrThrGlyVal 394
Db 1352 CTATGGTCTGTCATCAAAATATCAACGATGAAAGAGTATTTGCTGCTGTATGGATC 1411
Qy 395 ThrSerMetThrPheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMet 414
Db 1412 AGTGTGTGTTCTTTGCGGAGTATGTCGGGTGATGCTGACTCTGACCCCGGTCGTC 1471
Qy 415 SerLeuLeuSerGlyLeuValSerGlnValLeuSerThrThrThrMetLysAsnLeuAsp 434
Db 1472 TGCATGCTGCGGCATCGCTCTCCAAATGTTTTCAGCAGTATTTGGGGAT 1528
Qy 435 IleSerArg-----ProAspLysLysSerLysGlnGlnAspSerThrThrPro 451
Db 1529 ATGAAAGGAAACCCCTGTCGGAGCAGCAGTATGAGTATGAGTATGAGTATGAGTATG 1588
Qy 452 IleLysIleGluValAlaSerGlyMetIle----- 461
Db 1589 GGAACTTGTATGACAGGCGAGTAAAGTGAAGGAGCATGTGACAGAGCAAGAGAACT 1648
Qy 462 -----LeuValMetAlaPhePheLeu 468
Db 1649 GAAGAGGCTTGGGCCCCAATCAAAAGCATTTGACCATGTCGATGCTCATGCTCTCTG 1708
Qy 469 IleThrThrThrPheHisSerThrThrValThrSerGluAlaYrSerProSerIle 488
Db 1709 ATGATGTTGCGGGTCCACTGACGTCGGGTCAACAGCAACGCTACTCCAGTCAAGTGTG 1768
Qy 489 ValLeuSerAlaArgGlyGlyValSerArgIlePheAspAspPheArgGluAla 508
Db 1769 GTCTTGTCTCTACATCATGATGTTACAGGAGTATATATGATGATTTAGAGAGCG 1828
Qy 509 TyrThrThrLeuArgHisAsnThrProGluAspAlaLysValMetSerThrThrAspThr 528
Db 1829 TACTTTTGGCTGAGACAAACACCGATGAACACGCGCGGTCTATGCTGCTGGGACTAC 1888
Qy 529 GlyThrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrThrAsn 548
Db 1889 GGCTATCATGTTGTCGTCATGCGCAACAGGACCACTCTGTTGGTATACACACCTGGAAC 1948
Qy 549 AsnThrHisLeuSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaThrGlu 568
Db 1949 AACAGCCATCGCACTGTCGGAAGAGTATGCTTCCAATGAACGCGCGCTATATAA 2008
Qy 569 IleMetArgGluLeuAspValSerThrValLeuValIlePheGlyGlyLeuThrGlyThr 588
Db 2009 ATCATGAGTCCCTTGAATGATGATGTTGTTGTTATTTTCGAGGAGTATGCTAT 2068
Qy 589 SerSerAspAspIleAsnLysPheLeuThrMetValArgIleGlyGlySerThrAspThr 608
Db 2069 TCCGGGAGCATATCAACAGTTCCTCTGGATGCTCAGGATAGCT-----GAA 2116
Qy 609 GlyLysHis-----IleLysGluAsnAspThrThrThrProThrGlyGluPheArg 625
Db 2117 GGGAGATCCCAAGACATCGGAGAGTATTTTACCCAGGAGGAGTATCCGA 2176
Qy 626 ValAspArgGluGlySerProValLeuLeuAsnCysLeuMetThrLysMetCysThrThr 645
Db 2177 GTAGACAAAGCTGGGTCTCTACTCTGTAACTGCCTTATGATATAAAATGTCATACTAC 2236

Qy 646 ArgPheGlyGlnValThrThrGluAlaLysArgProGlyPheAspArgValArgAsn 665
Db 2237 AGATTTGGAGAAATGACAGCTAGATTTTCGCACTCCCCAGGCTTTCACCGAACAGTAAT 2296
Qy 666 AlaGluLeuGlyAsnLysAspPheGluLeuAspValLeuGluGluGlyThrThrThrGlu 685
Db 2297 GCTGAGATTTGAAATAAGACATTAATTCAGCATTTGGAGGAAGCTTTTACATCAGAG 2356
Qy 686 HisThrLeuValArgIleThrLysValLysAspLeuAspAsnArg 700
Db 2357 CACTGGCTTGTACAGATATATAAGTGAAGACACTGACACACAGG 2401

RESULT 14

US-10-437-963-99904
; Sequence 99904, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbatuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 99904
; LENGTH: 2779
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_97672C.1
US-10-437-963-99904

Alignment Scores:

Pred. No.: 1,69e-220 Length: 2779
Score: 2237.00 Matches: 419
Percent Similarity: 75.49% Conservative: 120
Best Local Similarity: 58.68% Mismatches: 145
Query Match: 60.51% Indels: 30
DB: 17 Gaps: 7

US-10-028-384-12 (1-705) x US-10-437-963-99904 (1-2779)

Qy 7 LeuArgLeuSerThrGluLysGlnAspThrLeuLysLeuLeuLeuLeuSerMetAla 26
Db 172 CTGGGCTGAAGACGAGCAGAGAGCTCTCTCGGGTCTCGGGCTGGCGCTCATC 231
Qy 27 AlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValIleHis 46
Db 232 TACGTGCTGGGTTCTGCTGCTCGGCTCTTCTCGGCTGCTCGCTACGAGTCCATGATCCAC 291
Qy 47 GluPheAspProThrPheAsnThrArgThrArgPheLeuAlaGluGluGlyPheThr 66
Db 292 GAGTTCGATCCCTACTTCAACTACCGCACCCAGCTCTTCTCTCCGACCATGATTCAGC 351
Qy 67 LysPheHisAsnThrPheAspArgAlaThrThrProLeuGlyArgIleGlyGly 86
Db 352 GAGTTCGAACTGGTTCGATTTCCAGAGAGCTGGTACCGCTCGCGCGCTCGTGGCGGCG 411
Qy 87 ThrIleThrProGlyLeuMetIleThrSerAlaAlaIleThrHisValLeuHisPhePhe 106
Db 412 ACCCTTCCCGGCTCTATGTCACCGCGCGCTGCTCCACCGCTCTCTCGCGGCTC 471
Qy 107 HisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerPhe 126
Db 472 TCCCTCGCGCTCCACATCCGAGGCTTGGGTCTCTACCCGCCCTCTCTCGCGGCAAC 531

127 ThrSerIleValThrTyrLeuLeuThryesleuLeuysAspAlaGlyLeuLeu 146
1591 TTGCTTAGTATAGTGTGTACACTGCACCTGGGTGAGCATCTGAAGCTTACTCTCTCTCT 1650
487 SerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArg 506
1651 TCCATTGTCTTGGCGCAAGGGGTCAATAATAGGAGCGAGGGTCAATTATTGATGATTTCGT 1710
507 GluAlaTyrTyrTyrLeuArgHisAsnThrProGluAspAlaLysValMetSerTyrTrp 526
1711 GAAGGCTATTATGGCTTCGTTCAGAACACTCTCTCTGATGCAAGATTATGTCATGTGG 1770
527 AspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnThr 546
1771 GATTATGATACCAATACAGCTATGGGAACAAGACTGTATTGTGTATGATACCAACA 1830
547 TrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGlyAla 566
1831 TCGAATAATACATATACAGTACAGTTCGGGTGCAATGTCATCTCTATGAGATGAGCG 1890
567 TyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThr 586
1891 TATGAGATAATGCAGTCACTGATGTGAATTATGTCCTGTCATTGTCGTTGTCGTTACT 1950
587 GlyTyrSerSerAspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThr 606
1951 GGTATTCTTCAGATGACATTAAACAAGTCTTATGATGGTGGCGATTGGTGGGA--- 2007
607 AspThrGlyLysHisIleLysGluAsnAspTyrTyrTyrProThrGlyGluPheArgVal 626
2008 ---GTTTTCCTGTAATCAAGAGCTGATTACCTGTT---AATGGGAGTATCGTGC 2061
627 AspArgGluGlySerProValLeuLeuLeuLeuLeuMetTyrLysMetCysTyrTyrArg 646
2062 GACAGGGGCGACACCAAAATGTTGAATGTCCTTAATGTCACAGCTTTGTTTATCCCGT 2121
647 PheGlyGluValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAla 666
2122 TTTGGAGAACTGACACAGANATGGAACCTCCAGGATACGATCGAGTACGAGGAGTG 2181
667 GluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyrThrThrGluHis 686
2182 GAGATTGGCAACAAGACATTAAGCTTGAGTACTTGGAGGAGGCATTCCACACTTCGAAC 2241
687 TrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
2242 TGGATAGTGGCATATACAGGTCAAAACCCCGAGAGATAGG 2283

RESULT 15
US-10-028-384-7
Sequence 7, Application US/10028384
Publication No. US20030148285A1
GENERAL INFORMATION:
APPLICANT: COMPATIGENE INC.
APPLICANT: PERREAU, Claude
APPLICANT: MCBRIDE, Kevin
TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
FILE REFERENCE: 5600-74
CURRENT APPLICATION NUMBER: US/10/028,384
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patent in version 3.1
SEQ ID NO 7
LENGTH: 2417
TYPE: DNA
ORGANISM: Drosophila melanogaster
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: AF132552
DATABASE ENTRY DATE: 1999-04-27
RELEVANT RESIDUES: (1)..(2417)
US-10-028-384-7
Alignment Scores:

QY 467 PheLeuIleThrTyrThrPheHisSerThrTrpValThrSerGluAlaTyrSerSerPro 486
Db 1591 TTGCTTAGTATAGTGTGTACACTGCACCTGGGTGAGCATCTGAAGCTTACTCTCTCTCT 1650
QY 487 SerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArg 506
Db 1651 TCCATTGTCTTGGCGCAAGGGGTCAATAATAGGAGCGAGGGTCAATTATTGATGATTTCGT 1710
QY 507 GluAlaTyrTyrTyrLeuArgHisAsnThrProGluAspAlaLysValMetSerTyrTrp 526
Db 1711 GAAGGCTATTATGGCTTCGTTCAGAACACTCTCTCTGATGCAAGATTATGTCATGTGG 1770
QY 527 AspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnThr 546
Db 1771 GATTATGATACCAATACAGCTATGGGAACAAGACTGTATTGTGTATGATACCAACA 1830
QY 547 TrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGlyAla 566
Db 1831 TCGAATAATACATATACAGTACAGTTCGGGTGCAATGTCATCTCTATGAGATGAGCG 1890
QY 567 TyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThr 586
Db 1891 TATGAGATAATGCAGTCACTGATGTGAATTATGTCCTGTCATTGTCGTTGTCGTTACT 1950
QY 587 GlyTyrSerSerAspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThr 606
Db 1951 GGTATTCTTCAGATGACATTAAACAAGTCTTATGATGGTGGCGATTGGTGGGA--- 2007
QY 607 AspThrGlyLysHisIleLysGluAsnAspTyrTyrTyrProThrGlyGluPheArgVal 626
Db 2008 ---GTTTTCCTGTAATCAAGAGCTGATTACCTGTT---AATGGGAGTATCGTGC 2061
QY 627 AspArgGluGlySerProValLeuLeuLeuLeuLeuMetTyrLysMetCysTyrTyrArg 646
Db 2062 GACAGGGGCGACACCAAAATGTTGAATGTCCTTAATGTCACAGCTTTGTTTATCCCGT 2121
QY 647 PheGlyGluValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAla 666
Db 2122 TTTGGAGAACTGACACAGANATGGAACCTCCAGGATACGATCGAGTACGAGGAGTG 2181
QY 667 GluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyrThrThrGluHis 686
Db 2182 GAGATTGGCAACAAGACATTAAGCTTGAGTACTTGGAGGAGGCATTCCACACTTCGAAC 2241
QY 687 TrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
Db 2242 TGGATAGTGGCATATACAGGTCAAAACCCCGAGAGATAGG 2283

Pred. No.:	1,95e-214	Length:	2417
Score:	2177.50	Matches:	415
Percent Similarity:	72.92%	Conservative:	110
Best Local Similarity:	57.64%	Mismatches:	144
Query Match:	58.90%	Indels:	51
DB:	15	Gaps:	9

US-10-028-384-12 (1-705) x US-10-028-384-7 (1-2417)

QY	16	ThrLeuLeuLysLeuLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu	35
Db	119	AGCCTAATCACTTCGCCATCTCGCTAATCGCTGGCTGGCGGGATTTTCCTTCGCTC	178
QY	36	PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyPheAsnTyArg	55
Db	179	TTGGCGGTCATCGGTTTCGAGTCGATATCCAGAGTTTGATCCCGTGGTTCAACTACCGG	238
QY	56	ThrThrArgPheLeuAlaGluGluGlyPheTyLysPheHisAsnTrpPheAspArg	75
Db	239	GCCACCGGCTACATGGTCAGAAATGGTGGTACAACTTCCTCAACTGGTTGCACGAGCG	298
QY	76	AlaTrpTrpProLeuGluValGileIleGlyThrIleTyTrpGluGlyLeuMetIleThr	95
Db	299	GCATGTTATCCGCTCGGAGAGATTGGCGGTACCGTCTATCCCGGCGTGATGATTACG	358
QY	96	SerAlaAlaIleTyHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal	115
Db	359	TCCGGCGGAATCCATTGGCTGTGTCAGCTACTCAACATACCGTCCATATTCGTGACATC	418
QY	116	CysValPheLeuAlaProLeuPheSerPheThrSerIleValThrTyLeuLeuThr	135
Db	419	TGCGTGTTCTGGCGCCGATCTTCACTAGTCGCTGACCTCCATCTCCACCTACTCGTGACC	478
QY	136	LysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValPro	155
Db	479	AAGAGAGCTGTGGTCGCGGGCGCGGCTCTTCGCGCGCAGCTTCATCGCCATCGTGCCT	538
QY	156	GlyTyIleSerArgSerValAlaGlySerTyTrpAsnGluGlyIleAlaIlePheCys	175
Db	539	GGCTACATCAGTAGTCGGTGGCTGGATCGTACGATAACGAGGGGATTGCCATATCGCC	598
QY	176	MetLeuLeuThrTyTrpMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla	195
Db	599	CTGCAGTTTCACCTACTTCCTGGGTGGGCTCAGTGAAGCTGGATCGGTCTTGTCGC	658
QY	196	AlaLysCysAlaLeuAlaTyTrpPheTyMetValSerSerTrpGlyTyTrpValPheLeu	215
Db	659	GCCGACCGGCTTTGTCTCTACTCTTACATGTTGTCGCTGGGTTGGTACGTGTTATC	718
QY	216	IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg	235
Db	719	ATCAACCTGATACCCCTGCACGTCCTTCGTACTGCTCATATTGGCGAGGTACTCGCCGCT	778
QY	236	IleTyTrpValAlaTyCysThrValTyCysLeuGlyThrIleLeuSerArgGlnIleSer	255
Db	779	CTGTGACCACCTACAGACCTTCTCATCTGGGACTGCTGTTCTCCATGCAGATCCCC	838
QY	256	PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly	275
Db	839	TTCTGGGATTCCAACCGATACGCACCATGCAACATGGCTGGCTGGGAGTGTGTTGTG	898
QY	276	LeuCysGlnIleHisAlaPheValAspTyTrpLeuArgSerLysLeuAsnProGlnPhe	295
Db	899	CTCCTTATGGCCGTGGCCACCTTCGGGCATTTGCACTCCGTCGTTCTCGCAACGAGTTC	958
QY	296	GluValLeuPheArgSerValIleSerLeuValGlyPheValLeuLeuThrValGly---	314
Db	959	CGGAAGCTGTTC-----ATCGTCGGCGGATGTGCTGGTGGCGGTGGGCTC	1003
QY	315	-----AlaLeuLeuMetLeuThr-----GlyLysIleSerProTrpThrGlyArgPhe	330
Db	1004	TTTGTGCCCCGTGGTGGTGTCAACATCTGGGCGTGTGTCGCCCGCTGGAGTGACGCTTC	1063

us-10-028-384-12.rnpb

Thu Dec 16 16:25:00 2004

Db	2126	GGTCCATCTGGATATGATCGCACACGTAAACGCCGTCATTTGGGAATAAGGACTTCGATCTG	2185
Qy	676	AspValLeuGluGluGlyThrThrGluHisTrpLeuValArgIleTyrLysValIys	695
Db	2186	ACCTACCTGGAGGAGGCCTACACCACAGACACTGGCTTTGTCGATCTATAGGTGAAG	2245

Search completed: December 15, 2004, 21:20:06
 Job time : 981.605 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:52:55 ; Search time 5457.63 Seconds
(without alignments)
4707.175 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 3697

Sequence: 1 MTKGFLRLSYEKDITLKL.....HMLVRIYKVKLDNRLGSLRT 705

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO spoal/US10028384/runat_14122004_131519_8030/app_query.fasta_1.3740
-DB=EST -QFMT=fastap -SUFFIX=first -MINMATCH=0.1 -IOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=2xt -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn_1_13960@runat_14122004_131519_8030 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREAP=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST :
1: gb_est1 :
2: gb_est2 :
3: gb_hic :
4: gb_est3 :
5: gb_est4 :
6: gb_est5 :
7: gb_est6 :
8: gb_ges1 :
9: gb_ges2 :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3622	98.0	3802	3 AK030363	AK030363 Mus muscu
2	3620	97.9	2730	3 AK081547	AK081547 Mus muscu
3	3605	97.5	3643	3 AK077877	AK077877 Mus muscu
4	3600.5	97.4	2656	3 BC028897	BC028897 Mus muscu
5	3495	94.5	2118	9 AV418284	AV418284 Homo sapi
6	3428.5	92.7	2103	9 AV418286	AV418286 Mus muscu
7	3419.5	92.5	2727	3 AK084770	AK084770 Mus muscu
8	2674	72.3	2079	9 AK418285	AK418285 Pan trogl
9	2239.5	60.6	4513	3 BC044321	BC044321 Xenopus 1

10	2237	60.5	2709	3 AK018758	AK018758 Mus muscu
11	2201	59.5	2669	3 AK012153	AK012153 Mus muscu
12	2199	59.5	2419	3 CNS0ADYZ	EX816490 Arabidops
13	1894	51.2	1345	3 CR606556	CR606556 full-leng
14	1740	47.1	1145	5 BX446691	BX446691 BX446691
15	1706	46.1	1124	7 CK030365	CK030365 AGENCOURT
16	1671.5	45.2	2334	3 AY310157	AY310157 Rattus no
17	1643	44.4	978	7 CR413615	CR413615 CR413615
18	1539	41.6	1141	6 CD505084	CD505084 CDA71-G02
19	1514	41.0	960	7 CN062150	CN062150 Ag2_P35_L
20	1455	39.4	1019	5 BX404479	BX404479 BX404479
21	1447	39.1	919	7 CR439276	CR439276 CR439276
22	1444	39.1	918	5 BX369502	BX369502 BX369502
23	1433	38.8	1037	4 BM451400	BM451400 AGENCOURT
24	1418	38.4	849	7 CR409802	CR409802 CR409802
25	1414	38.2	863	5 BX741571	BX741571 BX741571
26	1404	38.0	1185	6 CD505085	CD505085 CDA71-G02
27	1392	37.7	888	6 CD243529	CD243529 AGENCOURT
28	1378.5	37.3	1086	5 BU174083	BU174083 AGENCOURT
29	1377	37.2	831	7 CO424066	CO424066 UT-M-HUO-
30	1376.5	37.2	893	5 BX327064	BX327064 BX327064
31	1356.5	36.7	867	5 BU253920	BU253920 603414953
32	1354	36.6	862	5 BX713937	BX713937 BX713937
33	1354	36.6	863	7 CR440471	CR440471 CR440471
34	1354	36.6	863	7 CR442100	CR442100 CR442100
35	1344	36.4	892	7 CK407659	CK407659 AUF-1flv
36	1339	36.2	1154	4 BM542576	BM542576 AGENCOURT
37	1335.5	36.1	907	6 CB589879	CB589879 AGENCOURT
38	1317.5	35.6	1017	5 BU458020	BU458020 603367101
39	1314.5	35.6	933	5 BU913073	BU913073 AGENCOURT
40	1313	35.5	860	6 CA453997	CA453997 AGENCOURT
41	1310.5	35.4	816	5 BU200678	BU200678 603101974
42	1310.5	35.4	915	5 BX431319	BX431319 BX431319
43	1310.5	35.4	935	6 CB208848	CB208848 AGENCOURT
44	1309.5	35.4	1065	4 BM553035	BM553035 AGENCOURT
45	1307	35.4	903	6 CA487501	CA487501 AGENCOURT

ALIGNMENTS

AK030363 3802 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male pituitary gland cDNA, RIKEN full-length
enriched library, clone:530402C17, product:OLIGOSACCHARYL
TRANSFERASE STT3, SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
homolog [Mus musculus], full insert sequence.

ACCESSION AK030363.1 GI:26326354
VERSION AK030363.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

REFERENCE 2

10349636

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, N., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

REFERENCE 3

11042159

AUTHORS

Shibata, K., Itoh, N., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, T., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 Riken integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 MEDLINE
 PUBMED
 REFERENCE
 4

The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5

The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 3802)

Adachi, J., Aizawa, K., Akimura, T., Arai, K., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numata, K., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

Location/Qualifiers

1..3802
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:5330402C17"
 /db_xref="taxon:10090"
 /clone="5330402C17"
 /sex="male"
 /tissue_type="pituitary gland"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 133..2250
 /notes="unnamed protein product; OLIGOSACCHARYL TRANSFERASE
 STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
 homolog [Mus musculus] (SWISSPROT|P46978, evidence: PASTY,
 99.7%ID, 100%length, match=2115)
 putative"
 /codon_start=1
 /protein_id="BAC26921.1"
 /db_xref="GI:26326355"
 /translations="MTKGLFLRLSVKQDTLLKILNSAAVLFSTRLFAVLFPSV
 IHFEFYNYRTRFLAEGRFYKFNWEDDRAVPLRIGITGYPLMITSAAIYHV
 LHFPHITDIRNVCFLAFSFTITVYHLTKELKADAGLLAAIAVPGYISR
 SVAGSYDNEGIAIFCMLLTYYWIKAVKGTSGIYMAAKALAYFYWVSWSGGVFLINL

CDS

IPLHLVLMLTGRPSHRIYAYCTVYCLGTLISMOISFVGFQDPVLSEHMAARAGVGL
 COIHFADVYLCCKINPQOPVLPESVLSIVGFLVLTGALLMTGKISPWTCRFYSIL
 DSIYAKNIPILASVSEHQPWTSSYFEDQLLVFPPVGLYCFNSLSDARFILMY
 GVTISYFAVWRMLVLPVAPVILSGISVQVYSYVIMKMLDLSREKDKSKKQDSTY
 PIKESVAGSMILVMAFLITITFHSWTSEYSSPSVLSARGGGGSRIFIDDFEKA
 YYLHNTPEKVMWMDYGYQITAMNRTILVNNNTWNTNTHISVGGAMASTEKA
 YEIMREDSVYLVIKGLTGYSDINKFLMVRIGGSTETGRHKENDYKPTGEF
 RVDREGSPVLLNCLMKCYRFGQVYTEAKRPFGPDRVFNASIGNKDFELDLREAY
 TTEHLVRIYKVKDLNRLSRT"

ORIGIN

Alignment Scores:
 Pred. No.: 0 Length: 3802
 Score: 3622.00 Matches: 692
 Percent Similarity: 98.72% Conservative: 4
 Best Local Similarity: 98.16% Mismatches: 9
 Query Match: 97.97% Indels: 0
 DB: 3 Gaps: 0

US-10-028-384-12 (1-705) x AK030363 (1-3802)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluYsGlnAspThrLeuLeuLysLeu 20
 Db 133 ATGACTAAGCTTGGATTGCTTATGCTATGAGAGGAGGACACACTTCTAAAGCTT 192
 QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 Db 193 CTCATCCTGTCATGCTGCTGTGTATCTTCTACTGCTCTTCTGCTGCTGCTGAGA 252
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 Db 253 TTGGAAGTGTCCATCCATGAGTTGTCCTCGTACCTTAAATATCGGACTACCCGGTTCTG 312
 QY 61 AlaGluGluGlyPheTyrLysPheHisSerTrpPheAspArgAlaTyrTyrProLeu 80
 Db 313 GCTGAGGAGGGGTTTAAATTCATTAACCTGGTTGATGACCGGGCTTGGTACCTTTG 372
 QY 81 GlyArgIleLeuGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 373 GGCCGAATCAATGGAGGAAACAATTAACAGGTTTAATGATCACTTCTGCTGCAATCTAC 432
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db 433 CATGACTCCATTTCTCCATCATTGACATTGCAATCGGAATGCTGTGTTTCTTCTGGCC 492
 QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 Db 493 CGCTTTCTCCTCTTTCCACCACCATGTTACGTACCACTTACCAAGAGCTCAAGGAT 552
 QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
 Db 553 GCAGGAGCTGGGCTTCTTCTGCTGCCATGATTCTGTAGTTCCTGGTATATTTCTCGA 612
 QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 Db 613 TCTGTAGCTGGCTCTCTATGATAATGAAGAAATGCTATCTTTGCACTGCTTACTTAC 672
 QY 181 TyrMetTrpIleLeuAlaValLysThrClySerIleCysTrpAlaAlaLysCysAlaLeu 200
 Db 673 TACATGTGATCAAGGAGTGAAGACTGGTTCATCTATTGGGGCTGCCAAGTGTGCCCTC 732
 QY 201 AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuLeuAsnLeuLeuPro 220
 Db 733 GCTTATTCTACATGGTCTCTTCTATGGGAGGCTATGTGTTCTCTGATCAACTGATTCCT 792
 QY 221 LeuHisValLeuValLeuMetLeuThrClyArgPheSerHisArgIleTyrValAlaTyr 240
 Db 793 CTACATGCTCGTGTCTTAATGCTCAGAGGCCGCTTTTCTCACCAGATCTACGTAGCTAC 852
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db 853 TGTACTGTTTACTGCTGGGACCACTTCTTCTATGAGAGATTTCTTGTGTGTTGTTCCAG 912

QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuGlyCysGlnIleHis 280
 DB 913 CCGGTCCTTTTCATCAGAACACATGGCAGCCTTTGGAGTGTTCCTGTCTGTGATCCAT 972
 QY 281 AlaPheValLeuPheValLeuSerGlyLeuAsnProGlnGlnPheGluValLeuPheArg 300
 DB 973 CTTTCGTAGATACCTGTGCGACAGTGTGAATCCACAGCAATTCGAAGTCTCTTTCGG 1032
 QY 301 SerValLeuSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 DB 1033 AGTGTATCTCCCTGGTGGCTTTGTCTCTCTCCTCCTCCTGCTGGAGCTCTCCTCATGTAA 1092
 QY 321 GlyValLeuSerProTTPThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 DB 1093 GGAATAATTTCTCCCTGACAGGGCGTTTCTCTCTCTGCTGGATCCCTCTTATGCTAAG 1152
 QY 341 AsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 DB 1153 AATAACATTTCCCATTTATGTCATCTGTTTCTGAGCACCAGCCACAACTGGTCTTCTCTAC 1212
 QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 DB 1213 TATTTTGTATCAGAGCTCTCTGCTTCATGTTTCCAGTTGCGCTCATATTACTGCTTTAGC 1272
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1273 AACCTGTCTGATGCTCGGATTTTATCATCATGATGATGCTGACACAGCATGTACTTTTCA 1332
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1333 CTTGTATGTCGCTCTAAGCTGGTATTTGGACCTGTTATGTCATTTCTTCTGTCAT 1392
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLeuAsnLeuAspIleSerArgProAspLys 440
 DB 1393 GGTGTTTCCAGGTGGTGTCCACATATATGAAATCTGGACATAAGTCCGCCACAGCAAG 1452
 QY 441 LysSerIleLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 DB 1453 AAGAGCAAGACACAGGATTCATCTACCTTATTAAGATAGGTGGCGATGGGATG 1512
 QY 461 IleLeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 DB 1513 ATACTGGTCATGGCTTTCTTCTCATCACCTACACGTTTCATTCGACTGGGTGACCAAT 1572
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyValAspGlySerArgIle 500
 DB 1573 GAAGCCTATTTCTCCCTCATTTCTGCTCTGCTGGTGGGATGGCAGTAGGATC 1632
 QY 501 IlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 DB 1633 ATTTTGTATGACTTCCGAGAGCGGTATTAATGCTCGTCCGTCACATACTCCAGAGGATGCA 1692
 QY 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 DB 1693 AAGTCATGTCATGGTGGATATGGCTACCAATTAATCTCATGTCATGTCATGCAATTCGCAAT 1752
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1753 TTAGTGGACAATAACACATGAAATAATACCATATTTCTCGAGTAGGCGAGGCAATGGCA 1812
 QY 561 SerThrGluGlyLysAlaTyrGluIleMetArgLeuLeuAspValSerTyrValLeuVal 580
 DB 1813 TCCACAGAAGAAAGCCATGAAATCATGAGGAGCTTGTATGCTAGCTATGCTGTGTC 1872
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTrpMetVal 600
 DB 1873 ATTTTGGAGGCCCTTACTGGGTATTTCTGGATGATATCAACAAAGTTCTTTTGGATGTC 1932
 QY 601 ArgIleGlyGlySerThrAspThrGlyHisIleLysGlnAsnAspTyrTyrThrPro 620
 DB 1933 CGATTGGAGGAGCAGCAGAGGAGACACATTAAGGAGATGACTACTATCTCT 1992
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640

DB 1993 ACTGGGGAATTCGGTGTTCATGCTGAGGTTCTCGGTCTCTCAACTGCCTTATGAC 2052
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 DB 2053 ARAATGTGTACTACCGCTTTGGCAGGTCTACACAGAGCCAGCGTCCACAGGCTT 2112
 QY 661 AspArgValArgAsnAlaGlnIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
 DB 2113 GACCGTGTTCGAATCTGAGATTTGTAATAAGACTTTGAGCTTGATGCTCTGGAGAA 2172
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 DB 2173 GCCTATACACAGACACACTGGCTAGTCAGATATACAGGTAAAGACCTGATATCGA 2232
 QY 701 GlyLeuSerArgThr 705
 DB 2233 GGCTTGTCAAGGACA 2247

RESULT 2
 LOCUS AK081547
 DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:cl30038c18 product:OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG (BS) (INTEGRAL MEMBRANE PROTEIN 1) homolog (Mus musculus), full insert sequence.

ACCESSION AK081547
 VERSION AK081547.1 GI:26100006
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 PUBMED 10349636
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 PUBMED 11042159
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 PUBMED 11076861
 4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (Bases 1 to 2730)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,

Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, D., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
Location/Qualifiers
1..2730
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:Cl30038C18"
/db_xref="taxon:10090"
/clone="Cl30038C18"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
125..2243
/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus] (SWISSPROT|P46978, evidence: FASTA, 99.7%id, 100%length, match=2115)
putative"
2717..2722
/note="putative"
2730
/note="putative"

misc_feature
polya_signal
polya_site
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 2730
Score: 3620.00 Matches: 694
Percent Similarity: 98.73% Conservative: 3
Best Local Similarity: 98.30% Mismatches: 8
Query Match: 97.92% Indels: 1
DB: 3 Gaps: 0

US-10-028-384-12 (1-705) x AK081547 (1-2730)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
DB 125 ATGACTAAGCTTGGAATTTGGATTGCTCTATGAGACGACGACACTCTTAAGGTT 184
QY 21 LeuileLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 185 CTCACTCGTCGATGCTGCTGTTATCTTTTCTACTCGTCTTTTGTGCTGCTGAGA 244
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
DB 245 TTTGAAAGTGTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 304
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspAlaTyrTyrProLeu 80

DB 305 CTGAGGAGGGGTTTATTAATTCATTAACCTGGTTTCATGACCGGGCTTGTTACCTTTG 364
QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 365 GGCCGAATCATTTGGAGGAACAATTTACCCAGGTTTAATGATCACTCTCTGTCGAATCTAC 424
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
DB 425 CATGTACTCCATTTCTTCTTCATATCACCATTGACATTCGGATGCTCTGTGTTCTCTGCC 484
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
DB 485 CGGCTTTTCTCTCTTTTCCACCACCATGTTACGTACCCACCTTACCAAGAGCTCAAGGAT 544
QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
DB 545 GCAGAGCTGGGCTTCTTGTCTGCTGCCAATGATGCTGTAGTTCCTGGGTATATTTCTCGA 604
QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
DB 605 TCTGTAGCTGGCTCTCTATGATAATGAGGAATTCATCTTTTGCATGCTGCTTACTTAC 664
QY 181 TyrMetTyrIleLysAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaLeu 200
DB 665 TACATGTGGATCAAGGCAGTGAAGCTGTTCCATCTATTGGGCTGCCAAGTGTGCCCTC 724
QY 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIleAsnLeuIlePro 220
DB 725 GATTATTTCTACATGCTCTTTCATGGGAGGCTATGTTCTTCTGATCACTGATTCCT 784
QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
DB 785 CTACATGCTCTGCTGCTAATGCTGACAGCGCGTTTCTCACCGGATCTACGTAGCTTAC 844
QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
DB 845 TGTACTGTTTACTGCTGGGACCACTTCTTCTATGACAGATTCTCTTGTGTTGTTTCAG 904
QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
DB 905 CCGCTCTTTCTATCAGAACACATGCGAGCCTTTGGAGTGTGTTGTTCTCTGTCAGATCCT 964
QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
DB 965 GCTTTCGTAGATTACCTGCGCAGCAAGTTGAATCCACAGCAATTCGAAGTCTTTTCGG 1024
QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
DB 1025 AGTGTATTCTCCCTGGTTGGCTTTGTCTCTCTCACTGTGGAGCTCTCTCATGTAAACA 1084
QY 321 GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
DB 1085 GGAAAAATTTCTCCCTGGACAGGGCGTTTCTACTCTCTGCTGGATCCCTCTTATGCTAA 1144
QY 340 sAsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTyrSerTyr 360
DB 1145 GAATAACATTCCTATTATTCATCTGTTCTGAGCACCAGCCACACCTGCTCTTCTTA 1204
QY 360 rTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSe 380
DB 1205 CTATTTTGTATACAGCTCTTGTCTTTCATGTTTCCAGTTGGCCCTCTATTACTGCTTAG 1264
QY 380 rAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSe 400
DB 1265 CAACCTGTCTGATGCTCGGATTTTATCATCATGATGTTGTTGACGACCATGTACTTTC 1324
QY 400 rAlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyL 420
DB 1325 AGCTGTAATGGTGGCTTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1384
QY 420 eGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspL 440

QY 201 AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuLeuLeuLeuPro 220
 Db 620 GCCTTAATTCACATGCTCTCTTCATGGGAGGCTATGTTCTCTGATCAACTGATTCCT 679
 QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 Db 680 CTACATGCTCTGCTGTAATGCTGACAGCCGCTTTTCTCACCAGATATCATGATGCTTAC 739
 QY 241 CysThrValTyrCysLeuGlyThrLeuLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db 740 TGTACTGTTTACTGCTGGCACCATTCTTCTATCGCAATTCCTTGTGTTGTTCCAG 799
 QY 261 ProValLeuSerSerGlnHisValAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 Db 800 CCGCTGCTTTTCATCAGAACACATGGCAGCCTTTGGAGTGTGTTGGTCTCTGTCAGATCCAT 859
 QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db 860 GCTTTCGTAGATTACCTGGCAGCAAGTTGAATCCACAGCAATTCGAAGTCTTTTCGG 919
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 Db 920 AGTGTTATCTCCCTGCTGCTGTTGCTCTCCCTCATCTGTTGGAGCTCTCTCATGCTAACA 979
 QY 321 GlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 Db 980 GGAAATAATTTCTCCCTGGACAGGGCGTTTCTACTCTCTGCTGGATCCCTCTTATGCTAAG 1039
 QY 341 AsnAsnIleProIleLeuAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 Db 1040 AATAACATTCCTCCATTTATGCTATGTTCTGAGCACCAGCCACCACTGCTGTTCTGCTAC 1099
 QY 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 Db 1100 TATTTTGATCTACAGCTCCTCTGCTCATGTTTCCAGTTGGCTCTATTAATGCTTTAGC 1159
 QY 381 AsnLeuSerAspAlaArgIlePheIleLeuMetTyrGlyValThrSerMetTyrPheSer 400
 Db 1160 AACCTGCTCATGCTCGGATTTTATCATCATGTATGTTGACCATGATGATCTTTTCA 1219
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 Db 1220 GCTGTAATGCTGCTCTAATGCTGGTATGGCACCTGTTATGTCATCTTTCTGGCAT 1279
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 Db 1280 GGTGTTTCCAGTGTGCTCCCATATG-AAAAATCTGGACATAGTCGCCCCAGACAAG 1337
 QY 441 LysSerLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 Db 1338 AAGAGCAAGAGCAACAGGATCTACTTACCTTATTAAGATGAGTGGCGAGTGGGATG 1397
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 Db 1398 ATACTGCTCATGCTCTTTTCTCATCATCATGATGATGATGATGATGATGATGATGATGAT 1457
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 Db 1458 GAAGCTTATCTCTCTCCCTCATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1517
 QY 501 IlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 Db 1518 ATTTTGTAGACTTCGGAAGAGGTATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1577
 QY 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 Db 1578 AAAGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1637
 QY 541 LeuValAspAsnAsnThrTrpAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1638 TTAGTGGCAATAACACATGGAATAATACCCATATTTCTCGAGTAGGGCAGGCAATGGCA 1697

QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1698 TCCACAG 1757
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTyrMetVal 600
 Db 1758 ATTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1817
 QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1818 CGGATTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1877
 QY 621 ThrGlyGluPheArgValaAspArgGluGlySerProValLeuLeuLeuLeuMetTyr 640
 Db 1878 ACTGGGAG 1937
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 1938 AAAATGCTTACTACGCTTTGGCAGGCTCTACAGAGCAAGCGTCCACGAGCTTT 1997
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 1998 GACCGTGTTCGAAATGCTGAGATTGGTAATAAGACTTTGAGCTTGATGCTCTGGAGGAA 2057
 QY 681 GlyTyrThrThrGluHisTyrLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2058 GCGTATACACAGAACACTGGCTAGTCAGGATATACAGGTAAGGACCTGGGATATCGA 2117
 QY 701 GlyLeuSerArgThr 705
 Db 2118 GGCITGTCAAGGACA 2132

RESULT 5
 LOCUS AY418284 2118 bp DNA linear GSS 17-DEC-2003
 DEFINITION Homo sapiens ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION AY418284
 VERSION AY418284.1 GI:39774244
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2118)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 2118)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 source
 1. .2118
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 <1_>2118
 /gene="ITM1"
 /locus_tag="HCM6506"
 gene
 ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2118
 Score: 3495.00 Matches: 670
 Percent Similarity: 95.18% Conservative: 1
 Best Local Similarity: 95.04% Mismatches: 34
 Query Match: 94.54% Indels: 0
 DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x AV418284 (1-2118)

QY 1 MetThrIysPheGlyPheLeuArgLeuSerTyrGluIysGlnAspThrLeuLeuIysLeu 20
 DB 1 ATGACTAAGTTGGATTTTTCGGATTGCTTATGAGAGGAGGACACATTTTGAAGCTT 60
 QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 DB 61 CTCAATCTGTCATGCGTCTGCTATTATNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 DB 121 NNN 180
 QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaIleTyrProLeu 80
 DB 181 GCTGAGGAGGGTTTATAAATCCATAAAGTGGTTGATGACCGAGGCTGGTACCCTTG 240
 QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100
 DB 241 GGACGAATCATGGAGGAACAAATTTACCCAGGTTTAAATGATCACTCTGCTGCAATCTAC 300
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 DB 301 CATGTACTCCATTTTCCACATCACCATGACATTCGGAATGCTGTGTGTTCTCTGGGC 360
 QY 121 ProLeuPheSerPheThrSerIleValThrTyrLeuLeuThrIysGluLeuIysAsp 140
 DB 361 CTTCT 420
 QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
 DB 421 GCAGGGCTGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
 QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 DB 481 TCTGTGGCTGGCTCTCTATGAATAAGAGGATGCCATCTTTTGCAATGCTACTCACTAC 540
 QY 181 TyrMetTyrIleLysAlaValIysThrGlySerIleCysTyrAlaAlaLysCysAlaLeu 200
 DB 541 TACATGTGGATCAAGCGAGTAAGACTGGTTCCATCTGTTGGGAGCTAAGTGTGCCCTT 600
 QY 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIleAsnLeuIlePro 220
 DB 601 GCTTATTTCTCATGCTGCTCATGAGGAGGTTATGTTCTCTGATCAACTTAATTCCT 660
 QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 DB 661 CTCCAGCTCTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 DB 721 TGTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 DB 781 CTTGTCTTTTCTCATGAGCACATGCGAGCTTTTGGGCTTTTGGTCTCTCCAGATCCAT 840
 QY 281 AlaPheValAspTyrLeuArgSerIysLeuAsnProGlnPheGluValLeuPheArg 300
 DB 841 GCCTTTGTGGATTACCTCGCAGCAAGTTGAATCCACAACAAATTTGAAGTCTCTTTCCCG 900
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 DB 901 AGCGTCACTCTCTGCTAGGCTTTGCTCTCTCACTGCGGAGCTCTCTCTNNNNNNNN 960

QY 321 GlyIysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaIys 340
 DB 961 NGAAAAATATCTCCCTGGACGGGCGTTTCTACTCGCTGCTGGATCCCTCTTATGCTAAG 1020
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 DB 1021 AACAACTCCCATCATTTGCTGCTGAGCATCAGCCCAACCTGGTCTCTCATAC 1080
 QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValClyLeuTyrTyrCysPheSer 380
 DB 1081 TATTTTACCTTCGAGCTCTCGCTCTCATGTTTCCAGTTGGCCTCTATTACTGCTTTAGC 1140
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1141 AACCTGTCTGATGCCCGATTTTATCATCATGATGTTGGTGACCAAGCATGACTTTTCA 1200
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1201 GCTGTATGTTGGTGGCTCTAATGCTAGTGTGGACCTGTTATGTGCATTTCTCTGCAAT 1260
 QY 421 GlyValSerGlnValLeuSerThrTyrMetIysAsnLeuAspIleSerArgProAspLys 440
 DB 1261 GGAGTCTCCAGGTGCTGCTCCACATACATGAAGAACTGACATAAGTCTCCAGACAAG 1320
 QY 441 LysSerIysLysGlnGlnAspSerThrTyrProIleIysIleGluValAlaSerGlyMet 460
 DB 1321 AAGAGCAAGAGCAACAGGATTCACCTACCTATTAAAGAAATGAAGTGGAGTGGATG 1380
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 DB 1381 ATACTGTGCTGCTGCTTCTTCTCATCCTACCTACCTTTCATTCACCTGGGTGACCACT 1440
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyClyAspGlySerArgIle 500
 DB 1441 GAGGCTTACTCTTCTCGCTCCATGTACTATCTGCCGCTGGTGGGATGCGAGTAGATC 1500
 QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 DB 1501 ATATTTGATGACTTCCGAGAAGCATATTATTTGGCTTCGTCATAATACTCCAGAGGATGC 1560
 QY 521 LysValMetSerTrpTrpAspTyrGlyThrGlnIleThrAlaMetAlaAsnArgThrIle 540
 DB 1561 AAGTCTGATGCTCTGCTGGGATTTATGGCTATCAGATTACAGTATGCAACCAACCAAT 1620
 QY 541 LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1621 TTAGTGGACAATAACACATGGAATAATACCATATTTCTCGAGTAGGGCAGGCAATGGC 1680
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1681 TCCACAGAGGAAAAGGCTATGAGATCATGAGGAGCTCGATGTCAGTATGTCGCTGCTG 1740
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 DB 1741 ATTTTGGAGGCTCTGCTGGTATTCCTCTGATGATATCAACAAGTTTCTTTGGATGGTC 1800
 QY 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGlnAsnAspTyrTyrThrPro 620
 DB 1801 CGGATTGGAGGAGACAGATACAGGCAACATATCAAGGAGAATGACTATTATATCTCCA 1860
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 DB 1861 ACTGGGAGTTCCGTGGAGCGTGAAGGTTCTCCAGTGTGCTCAACTGCTCATGTAC 1920
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProArgIlePhe 660
 DB 1921 AAGATGTGTACTATCTGCTTGGACAGGTTTACAGANNNNNNNAGCGTCTCTCCAGGCTTT 1980
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnIysAspPheGluLeuAspValLeuGluGlu 680
 DB 1981 GACCGTGTCCGAATGCTGAGATTGGGAATAAGACTTTTGAGCTTGTGCTCTCTGGAGGA 2040


```

Db      1908  GGATGGTCCGATTGGAGGAGCAGACAGACAGGACACACATTAAAGGACATGACTACT 1967
Qy      618   yThrPrThrClyClyPheArgValArgGlyGlySerProValLeuLeuAsnCysL 638
Db      1968  ATACTCTTACCTGGGAATCCGTTGATCGTGAGGGTTCTCCGGTCTCTCACTGCC 2027
Qy      638  eMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysAspLeuA 658
Db      2028  TTATGTACAAATGGTTACTACCGCTTTGGCAGGCTCTACACAGAGCAAGCGCTCCAC 2087
Qy      658  roGlyPheAspArgValArgAsnAlaGluLeuGlyAsnLysAspPheGluLeuAspVal 678
Db      2088  CAGGCTTTGACCGTTCGAATGCTGAGATTGGTAATAGACTTTGACCTTGATGTC 2147
Qy      678  euGluGluGlyTyrThrThrGluHisThrPleuValArgileTyrLysValLysAspLeuA 698
Db      2148  TGGAGGAAGCGGTATACACAGAACACTGGCTAGTCAGGATATACAGGTAAGGACCTGG 2207
Qy      698  spAsnArgGlyLeuSerArgThr 705
Db      2208  ATATCGAGGGTTGTCAAGGACA 2230

RESULT 8
LOCUS      AY418285
DEFINITION Pan troglodytes ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION  AY418285
VERSION     AY418285.1 GI:39774245
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes
REFERENCE   1 (bases 1 to 2079)
AUTHORS     Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
            Todd,M.A., Tannenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
            Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
            Science 302 (5652), 1960-1963 (2003)
JOURNAL     14671302
PUBMED      2 (bases 1 to 2079)
REFERENCE   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
            Todd,M.A., Tannenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
            Direct Submission
            Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     These sequences were made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES    Location/Qualifiers
            source
            1..2079
            /organism="Pan troglodytes"
            /mol_type="genomic DNA"
            /db_xref="taxon:9598"
            <1..>2079
            /gene="ITM1"
            /locus_tag="HCM6506"

gene

ORIGIN

Alignment Scores:
Pred. No.: 1.77e-284 Length: 2079
Score: 2674.00 Matches: 534
Percent Similarity: 77.20% Conservative: 1
Best Local Similarity: 77.06% Mismatches: 158
Query Match: 72.33% Indels: 0
DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x AY418285 (1-2079)

```

```

Qy      1   MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
Db      1   ATGACTAAGTTTGGATTTTTCGATTTGCTCTATAGAGACAGACACANNCTGAGACCTT 60
Qy      21   LeuileuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db      61   CTCATTCTGTCAATGGCTGCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
Qy      41   PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db      121  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180
Qy      61   AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrPyrProLeu 80
Db      181  GCTGNNNNGGGTTTATAAATCCATAACTGTTTGTATGACGACGACCTGGTACCCCTTG 240
Qy      81   GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db      241  GGACGAATCATGGAGGANNNTTACCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 300
Qy      101  HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db      301  CATGTACTCCATTTTTCACATCACCATCGCAATTCGGAATGCTTNTGTGTCTCTGGCC 360
Qy      121  ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db      361  CCTCTCTTCTCTCTTACCACCATCGTCACGTACACCTTACCAAGAGCTCAAGGAT 420
Qy      141  AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
Db      421  GCAGGGGCTGGCTTCTGTCTGCTGCATGATGCTGTAGTCTCTCGATATATCTCCCA 480
Qy      161  SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db      481  TCTGTGGCTGGCTCTCTATGATATAGAGGATTCGCCATCTTTTGCATCTACTACCTAC 540
Qy      191  TyrMetThrIleLysAlaValLysThrGlySerIleCysTyrAlaLysCysAlaLeu 200
Db      541  TACATGTGGATCAAGGAGTAAGACTGGTTCCATCTCTGTGGCAGCTAAGTGTGCCCT 600
Qy      201  AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIlePro 220
Db      601  GCTTATTTTACATGGTCTCGTCATGGGAGGTTATGTGTCTCTGATCACTTAATCTCT 660
Qy      221  LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db      661  CTCCAGCTCTCTGCTGCTGATGCTCACAGGCGGTTTCTCTCACCGGATCTATGTGGCTAC 720
Qy      241  CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db      721  TGTACTGTTTACTGCTGGGACATATATCTTCTATGCAGATCTCTTGTGGGTTTCCAG 780
Qy      261  ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db      781  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 840
Qy      281  AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArg 300
Db      841  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 900
Qy      301  SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db      901  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 960
Qy      321  GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db      961  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020
Qy      341  AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrIrpSerSerTyr 360
Db      1021  AACACATCCCATCATTTGCTTCTGTGCTGTGAGCATCAGCCCAACCTGGTCTCTCATAC 1080

```


361	TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer	380
1081	TATTTTGACCTNNAGCTCCCTCGCTCTTCATGTTTCAGATTGGCTCTATTACTGCTTTNGC	1140
381	AsnLeuSerAspAlaAgllePheIleIleMetTyrGlyValThrSerMetTyrPheSer	400
1141	AACCTGTCGTATGCCGGATTTTATCATCATGATGTTGGTGACACAGCATGTACTTTTCA	1200
401	AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle	420
1201	GCTGTGAATGNN	1260
421	GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys	440
1261	GGAGTCTCCCAGGTNTGTCCNNATACATGAAGAANNTCNACATAANNNNNCANNAAG	1320
441	LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet	460
1321	AAGAGCAAGAGNNNNNAGGATTCGCCNCCCCTATTAAAGAATGAAGTGGCAAGTGGGATG	1380
461	IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer	480
1381	ATACTGGTCATGCGTTTCTTCTCATCACTACACCTTTCATTCACTGGGTGACCCAGT	1440
481	GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle	500
1441	GAGCGCTACTCTTCTCCGTCCCAITGTACTATCTGCCCGTGGTGGGATGGCAGTAGGATC	1500
501	IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla	520
1501	ATAATTGATGACTCCGAGAGAGCATATTATTGGCTTCGTCACATATCTCCAGAGGATGCG	1560
521	LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle	540
1561	AAGGTCTATNNCCTGGTGGGATTTATGCTATCATCAGATTACAGCTATGGCAAAACCGAACAT	1620
541	LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla	560
1621	TTAGTGACAAATACACATGGAAATATACCCATATTCTCNNNNNNNNNNNGCAATGGCG	1680
561	SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal	580
1681	TCCACAGAGGAAAAAGCCTATGAGATCATGAGGAGAGCTCGATGTCAAGCTATGTCTGGTC	1740
581	IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal	600
1741	ATTTTGGAGCCCTACTGGGTATTCCTCTGATGATATCAACAGTTNCTNNNGATGGTC	1800
601	ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro	620
1801	CGGATTTGGAGGGAGNACAGANNNANGCAACATATCAAGAGAGATGACTATTATATCTCCA	1860
621	ThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeuMetTyr	640
1861	ACTGGGAGATTCCGTGTGNNCMTGAAGTTCTCCAGTGTCTCTCANNNNNNNNNNNNNN	1920
641	LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe	660
1921	NNAGCGTCTCCAGCGTTT	1980
661	AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu	680
1981	GACCTGTCGGAATGCTGAGATTGGGAATAAAGACTTTTGAGCTTGATGCTCTGGAGGAA	2040
681	GlyTyrThrThrGluHisTrpLeuValArgIleTyrLys	693
2041	GCATATACCACAGAACTGGTGGTCAGAGATATACAAG	2079

RESULT 9	LOCUS	DEFINITION	4513 bp	mRNA	linear	HTC 23-JAN-2003
BC044321	LOCUS	Xenopus laevis,	clone IMAGE:4694177,	mRNA.		
BC044321	DEFINITION	Xenopus laevis,	clone IMAGE:4694177,	mRNA.		
BC044321	ACCESSION	BC044321				

[illegible]

Alignment
Pred. No.:
Score:
Percent Si
Best Local
Query Matc
DB:

US-10-028-

Qy

Dib

· ၄၇

20

5

五

3

2. 1

1

6.

[REDACTED]

BC044321.1 GI:27882209
HXC.
Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.
1 (base 1 to 4513)
Klein, S. and Strausberg, R.
Direct Submission
Submitted (17-JAN-2003) National Institutes of Health, Xenopus Gene
Collection (XGC), National Institute of Child Health and Human
Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
20892-7510, USA
NIH-MGC Project
Contact: XGC help desk
Email: cgaabs@mail.nih.gov
Tissue Procurement: Dr. Igor David
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Institute for Systems Biology
<http://www.systemsbioology.org>
contact: amadan@systemsbioology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: XGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
This is IRAC Plate: 94 Row: 1 Column: 19
One clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis. Similarity but not identity to protein
This clone has the following problem: frame shifted.
Location/Qualifiers
1 4513
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone_image="IMAGE:4684177"
/issue_type="Embryo, stage 31/32, Xenopus"
/clone_lib="NICHD_XGC_Emb4"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

OCES:	9, 076-236	4513
	239,50	428
length:		
Marches:		
Conservative:	74, 31*	103
Mismatches:	59, 94*	147
Indels:	60, 58*	36
Gaps:	3	8

4-12 (1-705) x BC044321 (1-4513)

6	ThrLeuLeuLeuLeuLeuSerMetLaLaLaValLeuSerPheSerThrArgLeu	35
3	TCCTGCTCTCTCTCACTATTATCTCTGGCCCTGGCTGGCGGCTTCACCTCAAGGCTC	522
6	PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg	55
3	TTCCGTGTCTCCGCTTCGAGAGCATCATCCACGAGTTCGACCCCTGGTTAACTACAGA	592
6	ThrThrArgPheLeuAlaGluGluGlyPheTyrLeuPheHisAsnTyrPheAspAspArg	75
3	TCTACACATCACTAGCATCCCATGGCTTTTATGAATTTTAAATTCGTTTGATGAACGG	642
6	AlaTrrTyrProLeuGlyArgIleGlyThrLeuTyrProGlyLeuMetIleThr	95
3	GCATGTATCCACTGGGGGAATAGTAGTGGAACAGATACCTCGGATTAATGGGTACA	702
6	SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrLeuPheIleArgAsnVal	115

[REDACTED]

Db 703 GCTGGGCTCATCCATGGATTTAAACATGTTAAACATAACCGTTTCATATACAGAGATGA 762
 Qy 116 CysValPheLeuAlaProLeuPheSerSerPheThrSerIleValThrValLeuLeuThr 135
 Db 763 TGTGTGTTCTTAGCACAGATATTAGTGGCTTACATCAATCCACTTTCCTGCTCACT 822
 Qy 136 LysGluLeuLeuLysAspAlaGlyValGlyLeuLeuAlaMetIleAlaValPro 155
 Db 823 CGAGAACTGTGGACCGGAGCGGAGCTTCTAGCTCCCTGCTTCACTGCTAGTGCCA 882
 Qy 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 Db 883 GGATACATATCCAGATCAGTAGCTGGATCAATTCACAAATCAAGGCATGCTACTTTGCA 942
 Qy 176 MetLeuLeuThrTyrTyrMetTyrIleLysAlaValLysThrGlySerIleCysTyrAla 195
 Db 943 CTGCAGTTTACATATTATTTGGGTAAATCTGTAAACACAGCTCTGTCTCTGGCCA 1002
 Qy 196 AlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeu 215
 Db 1003 ATAGTTGTGTGCTTATTTTATATGTTTCCGCTGGGTGGTTATGTGTTATC 1062
 Qy 216 IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
 Db 1063 ATTAATCTTATTCACATACATGTTTGTACTGTACTGATGAGAGATACACAAGAGA 1122
 Qy 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
 Db 1123 GTCTACATAGCTTACAGCACATTTTACATGCTGGTTTAACTATCATATGCAGATTC 1182
 Qy 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
 Db 1183 TTGTGGGATTTACAGCAATACAGCAAGTGAACATATGCGAGCTGCGAGGGTCTTTGCC 1242
 Qy 276 LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPhe 295
 Db 1243 CTGCTGCAAGCTATGCTCTCTCAGTACTTTGAGACACAACTAAGTAAAGCAGGATT 1302
 Qy 296 GluValLeuPheArgSerValIleSerLeu---ValGlyPheValLeuLeuThrValGly 314
 Db 1303 CAGACACTGTTTTCCTCGGTGTATCGTGGACCTGGAGCTGATTTCTCACTGTC--- 1359
 Qy 315 AlaLeuLeuMetLeuThrGlyLysIleSerProThrThrGlyArgPheTyrSerLeuLeu 334
 Db 1360 ATCTATTTCATTTACAGGCTACATTCCTCCATGGAGTGGAGATTTTACTCATTGG 1419
 Qy 335 AspProSerTyrAlaLysAsnIleProIleIleAlaSerValSerGluHisGlnPro 354
 Db 1420 GACACAGGGTATGCGAAATCCATATCCCATCTGCTGCTGCTGCTGCTGCTGCTGCT 1479
 Qy 355 ThrThrTyrSerSerTyrTyrPheAspLeuGlnLeuLeuValPheMetPheProValGly 374
 Db 1480 ACAACATGGGTCTCTCTTTTGTCTGCTATATATAGTATGATCATTTCTGCGAGGA 1539
 Qy 375 LeuTyrTyrCysPheSerAsnLeuSerAspAlaGlyPheIleIleMetTyrGlyVal 394
 Db 1540 CTTTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1599
 Qy 395 ThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMet 414
 Db 1600 AGTGGCGTTTACTTGTGGGCTAATGTTGCTGTTTGTGCTGCTGCTGCTGCTGCTGCT 1659
 Qy 415 SerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAsp 434
 Db 1660 TGCATGCTGCTGCGCAATGCTCTTCCAAATGTTTTGAGCATTAATCTGCGGTGAT--- 1716
 Qy 435 IleSerArgPro----- 438
 Db 1717 ATGAAGCGGAAACCCCGGAGAGACAGCAGTACGAGAGAGCGAAATCTGTGATCA 1776
 Qy 439 -----AspLysSerLysGlnGlnAspSerThrTyrProIleLysIleGlu--- 455
 Db 1777 CTCTATGATAGCGCTGGCAATGTGAGAAAGATGTGTCGAGCAGGAGAAAGCAGAGGAA 1836

Qy 456 -----ValLaserGlyMetIleLeuValMetAlaPhePheLeuIleThr 470
 Db 1837 GGTCTTGGCCCTAATAAAGAGTATAGTCACTATCTTGTATGCTAATGCTGTGTATGATG 1896
 Qy 471 TyrThrPheHisSerThrTyrValThrSerGluAlaTyrSerSerProSerIleValLeu 490
 Db 1897 TTTCTCTACACTGTAGTGGTAACTAGCAATGCTTATCCAGCCCTAGTGTGATTTG 1956
 Qy 491 SerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyrTyr 510
 Db 1957 CCGTATACATCAATCAGATGACACGAAATATCTTGGATGACTTTCAGAGAAGCATACTAC 2016
 Qy 511 TrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrpAspTyrGlyTyr 530
 Db 2017 TGGCTAAGCAGATACAGATGAACATGCGCGTGTATGCTCTGCTGGGATATGGCTAC 2076
 Qy 531 GlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnLeuThr 550
 Db 2077 CAAATTCAGGAATGGCTAATCGAACCACTTTAGTAGACAACAATACGTGGCAACATAGC 2136
 Qy 551 HistSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMet 570
 Db 2137 CACATAGCTTGGTTGGAAAGCAATGTCTTCAATGAAACCGCTGCTTATGAAATATG 2196
 Qy 571 ArgGluLeuAspValSerTyrValLeuValIlePheGlyLeuThrGlyTyrSerSer 590
 Db 2197 AAAAGCTTGGATGTAGATTTATGTAATAATATTTGGGGGTGTAATTCGTATCTCGT 2256
 Qy 591 AspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThrAspThrGlyLys 610
 Db 2257 CATGACATCAACAATTCCTCTGGATGGTTCGAATAGCA-----GAAAGAGAA 2304
 Qy 611 His-----IleLysGluAsnAspTyrTyrThrProThrGlyGluPheArgValas 627
 Db 2305 CATCAAAAGATATACAGGAAGTGCATCTTACC---ACAAGGAGAGATTTCTGTATGA 2361
 Qy 627 PArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPh 647
 Db 2362 CAAGCTGGATCTCCACCTTGTCTCAATGCTTCATGATATAGATGTCATATACCGGTT 2421
 Qy 647 eGlyGlnValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaG 667
 Db 2422 TGGTGAATGAGCTTGAATTCGGAACACCCACCTGGATTTGACCGGACACGTAATGCTGA 2481
 Qy 667 uileGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyrThrThrGluHisThr 687
 Db 2482 AATTGGGATTAAGATTAATAATCAAGCATTTGGAGAGAGCATTTACATCGAGGACTG 2541
 Qy 687 pLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2542 GCTTGTAGATATACAAAGTTAAAAAGCTAGAAAAACAGA 2581

RESULT 10

AK018758 2709 bp mRNA linear HTC 03-APR-2004
 LOCUS AK018758
 DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:130006C19 product:hypothetical oligosaccharyl transferase (OTase) STT3 subunit containing protein, full insert sequence.
 ACCESSION AK018758
 VERSION AK018758.2 GI:26384577
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 REFERENCE Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636


```

QY 295 eGluValLeuPheArgSerValIleSerLeu---ValGlyPheValLeuLeuThrValGI 314
D 1073 CCAGACCCCTTTCTTTTGGGTGCTCTCACTAGCTGCAGGGCGTGTCTCTAGTGC-- 1130
QY 314 yAlaLeuLeuMetLeuThrGlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLe 334
D 1131 -ATCTATCTGCATACACAGCTTATATTGCACCATGGAGTGGCAGGTTTATTCACTATG 1189
QY 334 uAspProSerTyrAlaLysAsnIleProIleAlaSerValSerGluHisGlnPr 354
D 1190 GGATCTGGGTATGCAAAATACACATTCCTCAATATTGCATCATGCTGTCTGAACATCAGCC 1249
QY 354 oThrThrTrpSerSerTyrTyrPheAspLeuGlnLeuLeuValPheMetPheProValGI 374
D 1250 TACGACATGGGTGCTTCTCTCTTGTATCTACATATTCTGTATGTTCCACGAGG 1309
QY 374 yLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleLeuMetTyrGlyVa 394
D 1310 CCTATGGTCTGCTCAAAATATACAGATCAAGATGTAATTTGTCTGCTCTGTATCGCAT 1369
QY 394 lThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMe 414
D 1370 CAGTCTGCTCTACTTTGCGGAGTGTGTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1429
QY 414 tSerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAs 434
D 1430 CTGCATGCTGTGGCCATCGCTCTCTCAATGTTTTTGGACACTATTGGGGGT---GA 1486
QY 434 pIleSerArg-----ProAspLysSerLysGlnAspSerThrTyrPr 451
D 1487 CATGAAAGGGAACCCACCTGTCGAGGACGAGTGCATGAGTGCATGAGTGCATGAGTGCAT 1546
QY 451 oIleLysIleGluValAlaSerGlyMetIle----- 461
D 1547 AGGAAACTTGTATGACAAAGGACGAGTAAAGTGAGGAGCATGTGCAGAGCAAGAGAAACC 1606
QY 462 -----LeuValMetAlaPhePhe 468
D 1607 TGAAGAGGGCTTGGCCCAACATCAAAAGCATTTGTGACCATGCTGATGCTCATGCTCT 1666
QY 468 uIleThrTyrThrPheHisSerThrTyrValThrSerGluAlaTyrSerSerProSerIl 488
D 1667 GATGATGTTGGGGTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1726
QY 488 eValLeuSerAlaArgGlyLysArgIlePheAspAspPheArgGluAl 508
D 1727 GTCCTTCTGCTCTCAATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1786
QY 508 aTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrpAspTy 528
D 1787 GTACTTTTGGCTGAGACAAAACCGATGAAACACGCGCGGCTCATGCTGCTGCTGCTGCT 1846
QY 528 rGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAs 548
D 1847 CGGCTATCAGATGCTGGCTGGCCACAGGACCCACTCTGTGGATACACACCTGGAA 1906
QY 548 nAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGI 568
D 1907 CAACAGCCACATCGCACTGGTGGGAAAGCTATGCTTCCATGAAACGCGCGCTATAA 1966
QY 568 uIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyLeuThrGlyTy 588
D 1967 ATCATGAGTCTCTTGTATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2026
QY 588 rSerSerAspAlaIleAsnLysPheLeuTrpMetValArgIleGlySerThrAspTh 608
D 2027 TTCCGGGGACGATATCAACAGTCTCTCTGATGCTGATGCTGATGCTGATGCTGATGCTG 2074
QY 608 rGlyLysHis-----IleLysGluAsnAspTyrTyrThrProThrGlyGluPheAr 625
D 2075 AGGGAGCATCCCAAGACATCCGGAGAGGTGACTATTTCACCCAGGAGGAGATTCOG 2134
QY 625 gValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTy 645

```

2135 AGTAGAAGAGCTGGGTCTCTCTACTCTGTGTAAACTGCCCTATGTATAAATGTCATACTA 2194
 QY 645 rArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAspArgValArgAs 665
 D 2195 CAGATTTGGGAAATGACGCTAGATTTTCGCACTCCCGCAGGCTTTGACCGAACACGTA 2254
 QY 665 nAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGluGlyTyrThrTrGI 685
 D 2255 TCCTGATTTGGAAATGAAACATTAATTCACGATTTGGAGAAAGCTTTTACATCAGA 2314
 QY 685 uHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 D 2315 GCACCTGGCTGTGAGATATATAAAGTGAAGCAACGACCTGCACACAGG 2360

RESULT 12
 CENSADYZ
 LOCUS
 DEFINITION
 Arabidopsis thaliana Full-length cDNA Complete sequence from clone
 Arabidopsis thaliana (thale cress).
 ACCESSION
 VERSION
 KEYWORDS
 SOURCES
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
 1 (bases 1 to 2419)
 AUTHORS
 Castell, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C.,
 Menard, M., Cruaud, C., Quetier, F., Scarpelli, C., Schachter, V.,
 Temple, G., Caboche, M., Weissenbach, J., and Salanoubat, M.
 Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences:
 A Combined Approach to Evaluate and Improve Arabidopsis Genome
 Annotation
 JOURNAL
 REFERENCE
 2 (bases 1 to 2419)
 Genoscope.
 Direct Submission
 TITLE
 JOURNAL
 COMMENT
 Submitted (19-NOV-2003) Genoscope - Centre National de Sequencage :
 BP 191006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr
 The sequences are based on single pass reads.
 Life Technologies (a division of Invitrogen) members carried out
 full-length libraries construction : Temple G.
 Genoscope members carried out sequencing and annotation : Castelli
 V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,
 Schachter V., Weissenbach J., Salanoubat M.
 URV INRA : Clepet C., Caboche M.
 Annotation is based on the June 2003 version of the Arabidopsis
 genome released by MIPS (Munich Information center for Protein
 Sequences). 5 prime and 3 prime are assembled with Phrap.
 http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
 length
 http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
 FEATURES
 source
 Location/Qualifiers
 1..2419
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Col-0"
 /db_xref="taxon:3702"
 /clone="GSLTPGH51ZC09"
 /tissue_type="Hormone Treated Callus"
 /plasmid="pCMVSPORT_6"
 1..2419
 /gene="At1g34130"
 ORIGIN
 Alignment Scores:
 Pred. No.: 9, 5e-232 Length: 2419
 Score: 2199.00 Matches: 418
 Percent Similarity: 73.68% Conservativity: 111

QY 693 ThrThrGluHisTrpLeuValArgIleTyrValysValysAspLeuAspAsnArg 700
 Db 2256 ACAACATCCACTGGATGATCGTATTTACAGAGTTAACTCTCAATAGG 2309

RESULT 13
 CR606556 1345 bp mRNA linear HTC 21-JUL-2004
 LOCUS full-length cDNA clone CLOBB0092D10 of Neuroblastoma of Homo
 DEFINITION sapiens (human).

ACCESSION
 CR606556
 VERSION CR606556.1 GI:50487363
 KEYWORDS HTC; CNSLT cDNA
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
 REFERENCE 2 (bases 1 to 1345)
 GenScope.
 Direct Submission
 Submitted (20-JUL-2004) GenScope - Centre National de Sequencage :
 BP 131 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.

FEATURES
 source
 1. 1345
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CLOBB0092D10"
 /tissue_type="Neuroblastoma"
 /plasmid="pCMVSPORT_6"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,39e-198 Length: 1345
 Score: 1894.00 Matches: 360
 Percent Similarity: 99.17% Conservative: 0
 Best Local Similarity: 99.17% Mismatches: 3
 Query Match: 51.23% Indels: 0
 DB: 3 Gaps: 0

US-10-028-384-12 (1-705) x CR606556 (1-1345)

QY 343 IleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyrPhe 362
 Db 1 ATCCCCATATGCTCTCTGTGTGATGACATGACCCACAACTGGTCTCATATTTT 60

QY 363 AspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyxCysPheSerAsnLeu 382
 Db 61 GACCTGCAGCTCTGCTCTTCATGTTCCAGTTGGCCCTATATGCTTTAGCAACCTG 120

QY 383 SerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSerAlaVal 402
 Db 121 TCTGATGCCCGGATTTTATCATCATGTATGTGTGACCCAGCATGTACTTTTCAGCTGTA 180

QY 403 MetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerClyIleGlyVal 422
 Db 181 ATGTGCGCTTATGCTAGTGTGGACCTGTTATGTGCAATTTCTCTGGCAATGGAGTC 240

QY 423 SerGlnValLeuSerThrTyrMetTysAsnLeuAspIleSerArgProAspIlyssSer 442
 Db 241 TCCCAAGTCTGTCCACATACATGAAGATCTGGACATAGTCTGACCAAGAGAGC 300

QY 443 LysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMetIleLeu 462
 Db 301 AAGAAGCAACAGGATTCACCTACCTATTAAGAAATGAAGTGGCAAGTGGATGATGACTG 360

QY 463 ValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThrSerGluAla 482
 Db 361 GTCATGGCTTTCTTCTCATCACCTACACCTTCATTCACCTGGTGACCAAGTGGGCC 420

QY 483 TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
 Db 421 TACTCTTCTCGTCCATTTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480

QY 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysVal 522
 Db 481 GATGACTTCCGAGGAGCATATTTGGTCTCGTCATATATCTCCAGAGGATCCGAGGTC 540

QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
 Db 541 ATGTCCTGTGGGATTTATGGCTATCAGATTACAGCTATGGCAACCCGAACTTTTAGTG 600

QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
 Db 601 GACATAACACATGGAATAATACCATATTTCTCGATGAGGAGGCAATGCGCTCCACA 660

QY 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
 Db 661 GAGGAAAAAGGCTATGAGATCATGAGGAGGCTCGATGTGCTATGTGCTGTCATTTTT 720

QY 583 GlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArgIle 602
 Db 721 GGAGCCCTCACTGGGATTTCTCTGATGATATCAACAAGTTTCTTTGGATGGTCCGGATT 780

QY 603 GlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThrGly 622
 Db 781 GGAGGAGCAGACATACAGGCAACATATCAAGGAGATGACTATATATCTCCACCTGGG 840

QY 623 GluPheArgValAspArgGluGlySerProValLeuLeuLeuLeuLeuLeuLeuLeuLeu 642
 Db 841 GAGTTCCTGTGGACCGGTTCCTCCAGTGTCTCAACTGCTCACTCACTCACTCACTCACT 900

QY 643 CysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAspArg 662
 Db 901 TGTTACTATCGCTTGGACAGGTTTACAGAGCCAGCGTCTCCAGGCTTTGACCGT 960

QY 663 ValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyr 682
 Db 961 GTCCGAAATGCTGAGATTGGGATTAAGAGCTTTGAGCTTGTATGCTGAGGAGGACATAT 1020

QY 683 ThrThrGluHisTrpLeuValArgIleTyrLysValysAspLeuAspAsnArgGlyLeu 702
 Db 1021 ACCACAGAACATTTGGCTGTGTCAGGATATACAGTAAAGACCTGGATAATCGAGGCTTG 1080

QY 703 SerArgThr 705
 Db 1081 TCAAGGACA 1089

RESULT 14
 BX446691 1145 bp mRNA linear EST 06-MAY-2004
 LOCUS BX446691 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
 DEFINITION CLOBB0092D10 5-PRIME, mRNA sequence.

ACCESSION
 BX446691
 VERSION BX446691.2 GI:47064152
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)

COMMENT

On May 22, 2003 this sequence version replaced gi:31027756.

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.

This sequence belongs to sequence cluster 6310.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna/s-CL0BB009ZD10RPL&c=6310.f.

FEATURES

Location/Qualifiers

1..1145

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CL0BB009ZD10"

/tissue_type="NEUROBLASTOMA"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Alignment Scores:

Pred. No.: 2,15e-181 Length: 1145
Score: 1740.00 Matches: 343
Percent Similarity: 95.56% Conservative: 1
Best Local Similarity: 95.28% Mismatches: 14
Query Match: 47.07% Indels: 4
DB: 5 Gaps: 2

US-10-028-384-12 (1-705) x BX446691 (1-1145)

QY	523	MetSerTrpTrpAspTyrGlyThrAlaMetAlaSerThrIleuVal	542
DB	541	ATGTCTCTGGTGGATTATGGCTATCAGATTACGATATGGCAACCAATTTTAGTG	600
QY	543	AspAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr	562
DB	601	GACATACACATGATGAAATACCCATATTTCTCGATAGGCGAGCAATGCGCTCCACA	660
QY	563	GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe	582
DB	661	GAGGAAAAGCCCTATGAGATCATGAGGAGCTCGATGTCACTATGCTGGTCAATTTT	720
QY	583	GlyGlyLeuThrGlyTyrSerSerAspIleLeuLysPheLeuTrpMetValArgIle	602
DB	721	GGAGGCGCTCCTGGGTATTCCTCTCATGATATCAACAAGTTCTTTGGATGGTCCGATT	780
QY	603	GlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThrGly	622
DB	781	GGAGGAGCACAGATACAGGCAACATATCAAGGAGATGACTATTATCTCAACTGGG	840
QY	623	GluPheArgValAspArgGlySerProValLeuLeuAsnCysLeuMetTyrLysMet	642
DB	841	GAGTTCCGTTGGCGCGTGAAGTTCTCCAGTGTCTCAACTGTCTCATGTACAGAIG	900
QY	643	CysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPheAspArg	662
DB	901	TGTTACTATCGTTTGGACAGGTTTACACAGAGCAAGCGTCTCCAGGCTTTGACCGT	960
QY	663	ValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyr	682
DB	961	GTCCGAAATCGTGAATGGGAATAAGACTT-GAGCTTGATGTCTCTCGAGGAR---CAT	1016
QY	683	ThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArgGlyLeu	702
DB	1017	ATACACACATTTGGCKKACGRTATT---ANAGTAAGRCCTGGATAT-CGAGGCGCTG	1072

CK030365	1124 bp	mRNA	linear	EST 26-NOV-2003
AGENCOURT 16620304	NIH_ZGC_7	Danio rerio	cDNA clone	IMAGE:7052144
5', mRNA sequence.				
CK030365	1	GI:38556289		
EST.				
SOURCE	Danio rerio (zebrafish)			
ORGANISM	Danio rerio			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Actinopterygii; Neopterygii; Teleostei; Ostariophysi;			
	Cypriniformes; Cyprinidae; Danio.			
REFERENCE	1 (bases 1 to 1124)			
AUTHORS	NIH-MGC http://mgc.ncbi.nlm.nih.gov/			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics / NIH National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgapbs@nih.gov Tissue Procurement: Len Zon, Harvard cDNA Library Preparation: Open Biosystems cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1AM14825 row: m column: 06 High quality sequence stop: 853.			

FEATURES	Location/Qualifiers
source	1..1124
	/organism="Danio rerio"
	/mol_type="mRNA"
	/db_xref="taxon:7955"

/clone="IMAGE:7052144"
 /tissue type="whole body"
 /lab host="DH10B"
 /clone lib="NIH ZGC 7"
 /note="Vector: pXpress1; Site 1: NotI; Site 2: EcoRV;
 Bulk tissue was collected from a whole adult individual
 from the Tuebingen strain. 1st strand cDNA was primed with
 a Not I - oligo(dT) primer, double-stranded cDNA was
 cloned into the Not I and EcoRV sites of pXpress-1.
 Library was size-selected for >1 kb fragments and
 normalized. A non-normalized version of this library is
 also available (NIH ZGC 10). Library was constructed by
 Open Biosystems (Huntsville, AL)."

ORIGIN

Alignment Scores:

Pred. No.: 1,22e-177 Length: 1124
 Score: 1706.00 Matches: 337
 Percent Similarity: 94.84% Conservative: 12
 Best Local Similarity: 91.58% Mismatches: 17
 Query Match: 46.15% Indels: 4
 DB: 7 Gaps: 0

US-10-028-384-12 (1-705) x CK030365 (1-1124)

QY 340 LysAsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSerSer 359
 Db 1 AAAAACAACATTCCTCATCGCTCTGTGTCTGACATCAGCCACCATGGTCTCTCC 60
 QY 360 TyrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPhe 379
 Db 61 TACTACTTTGACCTCCAGCTCTGGTCTTCATGTTCCAGTTGGCCTTTACTACTGCTTC 120
 QY 380 SerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPhe 399
 Db 121 AACAACTCTCTGATGCCAGCATTTTCATCATGTATGGCGTCCACGACATGTACTTC 180
 QY 400 SerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGly 419
 Db 181 TCAGCTGTGATGTGGTCTCTCATGTTGGTCTCGCTCCAGTAATGTGCATCTCTGCTGGC 240
 QY 420 IleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAsp 439
 Db 241 ATTGGAGTCTCTCAGGTCTCCACCTTCATGAGAACCTGGATGTCTAGTGCAGCTGAC 300
 QY 440 LysLysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGly 459
 Db 301 AAGAAACCAAGAGCAGCAGGACTCCACTACCCCAATCAAGAACGAGTGCAGAGTGGG 360
 QY 460 MetIleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThr 479
 Db 361 ATGATTCTAGTATGGCGTTCTTCCTCATCATACATACATCTCCACTCCAGCTGGTGACC 420
 QY 480 SerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyValAspGlySerArg 499
 Db 421 AGCGAGGGGTATTCTGTCCTCCCTCCATCGTCTCTGCTCGGGTGGCGAGCCGCGC 480
 QY 500 IleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAsp 519
 Db 481 ATCATCTTTGATGACTTCAGAGAGGCTTATTACTGGCTCAGACACACACACACAGGAT 540
 QY 520 AlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThr 539
 Db 541 GCTAAAGTGTATGCTGGTGGGATTTATGGGTATCAGATAACGAGATGGCCCAATCGAACG 600
 QY 540 IleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMet 559
 Db 601 ATTCTGGTTGACAATAACACTTGAATAACACTCACATCTCCAGAGTGGGTGAGGCAATG 660
 QY 560 AlaSerThrGluGlnLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeu 579
 Db 661 GCCTCCACAGAGAGAACCTTATGAGATTATGCGAGAGCTGGATGTGAGCTACGCTCCTG 720

QY 580 ValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMet 599
 Db 721 GTGATCTTTGGGGATTGACCGGATATTTCATCTGATGATATTAATAGTTCTCTGGGATG 780
 QY 600 ValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThr 619
 Db 781 GTGCGTATTCGGTGGGAGCAGACACAGGAGGAGCACAATAAAGAGCAGCATTTATACACG 840
 QY 620 ProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMet 639
 Db 841 CCACCTGGAGAGTTTCGTGTGACCGCGAGGGCTCTCTCTGCTCAACTGGCTCATG 900
 QY 640 TyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGly 659
 Db 901 TACAAGATGTGCTACTATCGCTTCNGCCAGGTCTACACCGAAGCCAG-CGTCTCTCTGT- 958
 QY 660 PheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 679
 Db 959 TACGAAAGAGTTCTGTATGCTGAGATTGGGCAATAGGACTTTTAACTGGAGCTGCTGGAA 1018
 QY 680 GluGlyTyrThrThrGluHisTrpLeuValArgIle-TyrLysValLysAspLeuAspA 699
 Db 1019 AAAAGCTTTAAACTACAAAACACTGGCTGGCGCAGATAATACAGGCCAAGATTTTACA 1078
 QY 699 snAtgGlyLeuSerArgThr 705
 Db 1079 ATCGGGCCCTTCCGAAACC 1098

Search completed: December 15, 2004, 11:54:55

Job time : 5547.63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model.

Run on: December 14, 2004, 13:16:57; Search time 7810.5 Seconds
(without alignments)
4268.513 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 705

Sequence: 1 MTKFGFURLSYEKQDTLLKL.....HMLVRIYKYKLDNRGLSRT 705

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10028384/runat.14122004.131645.6800/app_query.fasta.1.3740
-DB=genEmbl -QWTF=fastcap -SUFFIX=Oligo.rge -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -WAITRX=Oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn.1.1.20262 @runat.14122004.131645.6800 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	705	100.0	2472	6	AX799092 Sequence
2	705	100.0	2472	9	HUMB5A
3	226	32.1	2118	9	BT007100 Homo sapi
4	226	32.1	2118	12	BT008132 Synthetic

SUMMARIES

RESULT 1
AX799092
LOCUS AX799092 2472 bp mRNA linear PAT 08-OCT-2003
DEFINITION Sequence 11 from Patent WO03054008.
ACCESSION AX799092
VERSION AX799092.1 GI:37605063
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Perreault, C. and McBride, K.
JOURNAL Mammalian SAMP protein, gene sequence and uses thereof in cancer therapy
Patent: WO 03054008-A 11 03-JUL-2003;
Compatisgene Inc. (CA)
FEATURES
Location/Qualifiers
source
1. .2472
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/note="ITM1, accession No NM-CO02219"

ALIGNMENTS

5	226	32.1	2329	6	CQ834708	Sequence
6	226	32.1	2339	6	CQ834706	Sequence
7	226	32.1	2435	6	CQ834710	Sequence
8	226	32.1	2450	9	BC048348	Homo sapi
9	226	32.1	2458	6	CQ834714	Sequence
10	226	32.1	2472	6	CQ727867	Sequence
11	226	32.1	2516	9	BC020965	Homo sapi
12	226	32.1	2760	6	CQ492624	Sequence
13	220	31.2	2898	9	HS0809254	Sequence
14	192	27.2	1796	9	AK130548	Homo sapi
15	186	26.4	2293	9	HUMTMC	Sequence
16	155	22.0	1094	5	CR386917	Callus ga
17	135	22.0	2567	5	BC063234	Danio rer
18	155	22.0	2608	5	BC046072	Sequence
19	152	21.6	2736	10	BC037612	Mus muscu
20	152	21.6	3094	6	AX799090	Sequence
21	152	21.6	3094	10	MUSTRPR	Sequence
22	146	20.7	440	6	AX351312	Sequence
23	145	20.6	2481	5	BC057313	Xenopus t
24	144	20.4	2470	6	CQ834712	Sequence
25	144	20.4	3666	10	AK129027	Sequence
26	140	19.9	455	6	AX341209	Sequence
27	128	18.2	487	6	BD229538	Human gen
28	128	14.5	414	6	AX884408	Sequence
29	102	14.5	414	6	BD024018	Sequence
30	102	14.5	487	6	BD273912	Sequence
31	102	14.5	487	6	AR277433	Sequence
32	102	14.5	487	6	AR407768	Sequence
33	102	14.5	487	6	AR441618	Sequence
34	102	14.5	487	6	AX368945	Sequence
35	101	14.3	307	6	AR272646	Sequence
36	101	14.3	307	6	AR273528	Sequence
37	101	14.3	307	6	AR276227	Sequence
38	101	14.3	307	6	AR277109	Sequence
39	101	14.3	307	6	AR406502	Sequence
40	101	14.3	307	6	AR407384	Sequence
41	101	14.3	307	6	AR440352	Sequence
42	101	14.3	307	6	AR441234	Sequence
43	101	14.3	307	6	AR472510	Sequence
44	101	14.3	307	6	AX062748	Sequence
45	101	14.3	307	6	AX367665	Sequence

Alignment Scores:
Pred. No.: 0
Score: 705.00
Length: 2472
Matches: 705

Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-10-028-384-12 (1-705) x AX799092 (1-2472)			
QY	1	MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu	20
DB	107	ATGACTAAGTTTGGATTTCCTGCGATTGCTCTATGAGAACGACGACACACTTTTGAAGCTT	166
QY	21	LeuLeuLeuSerMetAlaAlaValLeuLeuSerPheSerThrArgLeuPheAlaValLeuArg	40
DB	167	CTCATCTGCAATGGCTGCTGATATATCTTCACCTCGCTGTTGCTGCTCTGAGA	226
QY	41	PheGluSerValLeuHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu	60
DB	227	TTTGAAAGTGTATCCATGAGTTTGTATCGTACTTTTAATATCGGACTACCAAGGTTCCTG	286
QY	61	AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrTyrProLeu	80
DB	287	GCTGAGGAGGGGTTTATAAATTCCTAACTGGTTTGTATGACCGAGCCCTGGTACCCTTTG	346
QY	81	GlyArgIleLeuGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr	100
DB	347	GGACGAATCATTTGGAGGACAAATTTACCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC	406
QY	101	HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla	120
DB	407	CATGTACTCCATTTTCCACATCACCATCGCAITTCGGAAATGCTGTGTGTTCCTGGCC	466
QY	121	ProLeuPheSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp	140
DB	467	CCTCTCTTCCTTCCTTCCTCCATCCATCGTCACGTACCTCTCTTACCAAGAGAGCTCAAGGAT	526
QY	141	AlaGlyValGlyLeuLeuAlaAlaAlaMetIleAlaValProGlyTyrIleSerArg	160
DB	527	GCAGGGCTGGGCTTCTTGCTGCTGCGATGATTCGTAGTTCTCTGGATATATCTCCCGA	586
QY	161	SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr	180
DB	587	TCTGTGCTGGCTCCTATGATAATGAAGGATTCGCATCTTTTGCATGTACTACCTACCTAC	646
QY	181	TyrMetThrPheLysAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaLeu	200
DB	647	TACATGGAATCAAGCGATGAAGATGCTGCTTCATCTGTGGCAGCTAAGTGTGCCCTT	706
QY	201	AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuLeuLeuLeuPro	220
DB	707	GCTTATTTCTACATGGTCTCGTCAATGGGAGGTTATGTCTCTGATCAACTTAATTCCT	766
QY	221	LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr	240
DB	767	CTCCAGCTCCTGCTGCTGATGCTCACAGGCCGTTTCTCTACCGGATCTATGTGGCCCTAC	826
QY	241	CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln	260
DB	827	TGTACTGTTTACTGCTGGTACTATACCTTCTTAGGCAGATCTCTCTTTTGGGTTTCCAG	886
QY	261	ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis	280
DB	887	CCTGCTCTTTCATCAGACACATGAGGAGGTTTGGGCTTCTTGGTCTCTGCGCAGATCCAT	946
QY	281	AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGlnValLeuPheArg	300
DB	947	GCCTTTTCTGGATTACCTGCGCAGCAAGTTTGAATCCACAACTTTTGAAGTTCTTTTTCGG	1006
QY	301	SerValIleSerLeuValGlyPheValLeuLeuThrValGlyValLeuLeuMetLeuThr	320
DB	1007	AGCGTCATCTCTGTGTAGGCTTTGCTTCTCACCCTGGAGCTCTCTCTCATGTCTGACA	1066
QY	321	GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys	340

DB	1067	GGAAAAATATCTCCCTGGACGCGGGCGTTTCTACTCTGCTGGATCCCTCTTATGCTAAG	1126
QY	341	AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrThrTpSerSerTyr	360
DB	1127	RACACATCCCATCATTTGCTTCTGTCTGAGCATCAGCCACCACTCGTCTCTCATAC	1186
QY	361	TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer	380
DB	1187	TATTTTGGACCTGACCT	1246
QY	381	AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer	400
DB	1247	RACCTGTCTGATGCCCGGATTTTATCATCATGATGTTGACCATGATGATCTTTTCA	1306
QY	401	AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle	420
DB	1307	GCTGTAAATGGTGGCTTAATGCTAGTGTGGACCTGTTATGACCATTTCTCTGGCAT	1366
QY	421	GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys	440
DB	1367	GGAGTCTCCAGTCTCTCCACATACATGAAGATCTGGACATAAGTCCGCCACAGCAAG	1426
QY	441	LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet	460
DB	1427	AAAGCAAGAAACAAACAGGATTCACCTACCTACCTAATGAATTTGAAGTGGCAAGTGGGAT	1486
QY	461	IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTyrValThrSer	480
DB	1487	ATACTGGTCAATGGCTTCTTCTCATCACCTACACCTTTCATCACTGGGTGACCAT	1546
QY	481	GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle	500
DB	1547	GAGGCTACT	1606
QY	501	IlePheAspAspPheArgGluAlaTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr	520
DB	1607	ATATTGTGATCTCCGAGAGGACATATATTTGGCTCTGTCATATATCTCCAGAGATCG	1666
QY	521	LysValMetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle	540
DB	1667	AAAGTCAATGCTGCTGGGATTTATGGCTATCAGATTCAGCTATGGCAACCCGAAACAA	1726
QY	541	LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla	560
DB	1727	TGAGTGGACATTAACATGGAATATATACCATATTTCTCGAGTAGGCGAGCAATGGCG	1786
QY	561	SerThrGluGluLysAlaTyrGluLeuMetArgGluLeuAspValSerTyrValLeuVal	580
DB	1787	TCCACAGAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGTGCTGTC	1846
QY	581	IlePheGlyGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTyrMetVal	600
DB	1847	ATTTTGGAGGCTCTCTCTGGGTATTCCTCTGATGATCAACAAGTTTCTTGGATGTC	1906
QY	601	ArgIleGlyGlySerThrAspThrGlyLysIleLysGluLeuAsnAspTyrTyrThrPro	620
DB	1907	CGGATTTGGAGGAGACAGATACAGGCAACATATCAAGAGAGATGATGATTAATCTCCA	1966
QY	621	ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr	640
DB	1967	ACTGGGAGTTCCTGCTGGACCGTGAAGTTCTCCAGTGTGCTCTCACTGCTCATGTCAC	2026
QY	641	LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe	660
DB	2027	AAAGTGTGTACTATCTCTGCTTGGACAGGTTTACAGAGCAACAGCGCTCTCCAGGGCTTT	2086
QY	661	AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu	680
DB	2087	GACCGTCCGAAATGCTGAGATTTGGGAATAAGACTTTGAGCTTGTATGCTCTGGAGGAA	2146
QY	681	GlyTyrThrThrGluHisThrLeuValArgIleTyrLysValLysAspLeuAspAsnArg	700
DB	2147	GGCTATACACAGAACATTTGGCTGGTTCAGGATATACAGGTAAAGGACCTGGGATAATCGA	2206

```

QY 701 GlyLeuSerArgThr 705
Db 2207 GGC TTGTCAAGGACA 2221

RESULT 2
HUMBSA 2472 bp mRNA linear PRI 14-AUG-1996
LOCUS Human putative transmembrane protein precursor (B5) mRNA, complete
DEFINITION cds.
ACCESSION L38961
VERSION L38961.1 GI:624703
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2472)
AUTHORS Lissy,N.A., Bellacosa,A., Sonoda,G., Miller,P.D., Jhanwar,S.C. and
Testa,J.R.
TITLE Isolation, characterization, and mapping to human chromosome
11q24-25 of a cDNA encoding a highly conserved putative
transmembrane protein, TMC
JOURNAL Biochim. Biophys. Acta 1306 (2-3), 137-141 (1996)
MEDLINE 96221283
PUBMED 8634329
FEATURES
source
    1..2472
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    1..2472
    /gene="B5"
    1..106
    /gene="B5"
    107..2224
    /gene="B5"
    /note="putative transmembrane protein precursor"
    /codon_start=1
    /protein_id="AAB05994.1"
    /db_xref="GI:624704"
    /translation="MTKFGFLRLSYEKQDTLLKLLLSMAAVLSFTRLFAVLRFESV
    IHEPDYFNRTYTRFLAEQGYFHNWFDRAWYPLGRITGGTYVPLGMITGAAIVYH
    LHPFHITIDRNVCVFLAFLFSSFTSIVTYLLTKELKDAGLLAAAMTAVVPGVISR
    SVAGSYNEGIAFCMLTYVMKKAUKTGSICWAACKCALAVFYVSSWGMGVFVLNL
    IPLAVILMTGRFSHRIYVACTVCLTILSRQISFVCPVPLVSSERMGAFVFLGL
    CQIHAFVDYLRKLNPOQFEVLPFRSVLSLVGVLLTVGALLMTGKISPTWGRFYSLI
    DPTAKNNIPLIASVSHQPTWSSYIFDLQLVFPFPVGLYCFNSLSDARIFIMY
    GVTSMYSAVMVRMLVLPVNGILSGIVSQVLSYMNKLDISRDPKSKKQDQSTY
    PIKEVASGMILVMAFPLITYTFHSTWITSEAYSSPSIVLSARGDGSRIIPDDFREA
    YWLRHNTPEDAKYSWMDYGYQITAMANTILVDNNTWNTTHSRVGQAMASTEKA
    YEIMRELDVSYLVI FGLTGYSSDDINKFLMWVIRGGSTDGKHKENDYVTPTGEF
    RVDRGSPVLNCLMYKCYVREGQVYTEAKRPPGDFRVNAEIGNKDFELDLSEGY
    TTEHWLYRIYKVDLNRGLSRI"
    107..196
    /gene="B5"
    197..2221
    /gene="B5"
    /product="putative transmembrane protein"
    452..499
    /gene="B5"
    /note="encodes putative transmembrane domain"
    533..580
    /note="encodes putative transmembrane domain"
    /gene="B5"
    620..667
    /note="encodes putative transmembrane domain"
    /gene="B5"
    746..793
    /note="encodes putative transmembrane domain"
    /gene="B5"
    /note="encodes putative transmembrane domain"
    815..862
    /note="encodes putative transmembrane domain"
    /gene="B5"

```

```

/note="encodes putative transmembrane domain"
1013..1060
/gene="B5"
/note="encodes putative transmembrane domain"
1199..1246
/gene="B5"
/note="encodes putative transmembrane domain"
1310..1357
/gene="B5"
/note="encodes putative transmembrane domain"
1475..1522
/gene="B5"
/note="encodes putative transmembrane domain"
2223..2472
/gene="B5"
2455
/gene="B5"
3'UTR
polyA_site
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2472
Score: 705.00 Matches: 705
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-10-028-384-12 (1-705) x HUMBSA (1-2472)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 107 ATGACTAAGTTTGGATTTCCTTCATGAGAAGCAGGACACACTTTGAAGCTT 166
QY 21 LeuileLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGTCAATGGCTGCTGCTATTATCTCTTCCTCCTCGTCTGTTGCTGCTGAGA 226
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTGAAAGTGTATTCATGATGATTGATCCGTACTTTAATATCGGACTACCAAGGTTCCGTG 286
QY 61 AlaGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrPyrProLeu 80
Db 287 GCTGAGGAGGGGTTTATAAATCCATAACTGTTTGTATGACCGAGCTGTACCTTTG 346
QY 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100
Db 347 GGACGAATCATTTGGAGAACAAATTTACCCAGGTTTAATGATCACCCTCTCTCCTCAATCTAC 406
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTTCACATCATCCATCGACATTCGGAATGCTGTGTGTTCTGTGCC 466
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 526
QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
Db 527 GCAGGGGCTGGGCTTCTTGTGCTGCTGCCATGATTCGTGTAGTTCCTGTGATATATCTCCGA 586
QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGTGGCTGGCTCTCTATGATAATGAAGGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 646
QY 181 TyrMetThrIleLysAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGGATCAAGGACAGTAAAGACTGGTTCCTCATCTGTGGCAGCTAAGTGTGCCCTT 706
QY 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuLeuLeuLeuLeuLeu 220
Db 707 GCTTATTCTACATGTTCTGCTCATGGGAGGTATGTGTCTCTGATCAACTTAATCTCT 766

```

QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 DB 767 CTCACGTCCTCGTGCTGATGCTCACAGGCGGTTCTCTCACCGGATCTATGTCGCTAC 826
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 DB 827 TGTACTGTTACTGCTCGGCTACTATATCTTCTAGGACAGATCTCCTTTGTGGGTTCCAG 886
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 DB 887 CCTGCTCTTCATCAGACACATGGCAGGGTTTGGGGTCTTTGGTCTCTGCGCAGATCCAT 946
 QY 281 AlaPheValLeuSerLeuArgSerLeuSerLeuSerProGlnGlnPheGluValLeuPheArg 300
 DB 947 GCCTTTTGGGATTAACCTCGGCACAGATTTGAATCCCAACAATTTGAAGTTCTTTTCCGG 1006
 QY 301 SerValIleSerLeuValGlyPheValLeuThrValGlyValAlaLeuLeuMetLeuThr 320
 DB 1007 AGCGTCATCTCTCTGGTAGGCTTGTCTCTCACCGTGGAGCTCTCTCTCATGCTGACA 1066
 QY 321 GlyIleSerProThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaIys 340
 DB 1067 GGAATAATATCTCCCTCGAGCGGGCGTTCTACTCACTCTGTCATCCCTTTATGCTAAG 1126
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 DB 1127 AACACATCCCATCAATGCTTCTGTGTGACATCAGCCACACCACTGGTCTCTCATAC 1186
 QY 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 DB 1187 TATTTTACCTGACGCTCCCTCGCTTCTCATGTTTCCAGTTGGCTCTTATTTACTGCTTAGC 1246
 QY 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1247 AACCTGCTGATGCCCCGGAATTTTATCATCATGATGTTGTCACCATGATGACTTTTCA 1306
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1307 GCTGTATGTTGGTCTCAATGCTAGTGTGGACCTGTTATGAGCAATCTCTCTGGCAT 1366
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLeuAsnLeuAspIleSerArgProAspIys 440
 DB 1367 GGAGTCTCCAGGTGCTGTCCACATACATGAAGAATCTGGACATTAAGTCCGCCACAG 1426
 QY 441 LysSerIysGlnGlnAspSerThrTyrProIleIleGluValAlaSerGlyMet 460
 DB 1427 AAGAGCAAGAGCAACAGGATTCACCTACCTATTAAGATTGAAGTGGCAAGTGGGATG 1486
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 DB 1487 ATACTGGTCAATGCTTCTTCTCATCACTACCTACCTTCACTCACTGGTGACCAAGT 1546
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyAspIleSerArgIle 500
 DB 1547 GAGGCTACTCTTCTCGTCCATGTTACTATCTGCCCGTGGTGGGATGGCAGTAGGATC 1606
 QY 501 IlePheAspAspPheArgGluAlaTyrTyrLeuArgHisAsnThrProGluAspAla 520
 DB 1607 ATATTGATGACTTCCGAGAGCATAATTATGGCTTCGTCAATAACTCCAGAGGATGCG 1666
 QY 521 LysValMetSerTrpTrpAspTyrGlyThrGluIleThrAlaMetAlaAsnArgThrIle 540
 DB 1667 AAGGTCAATGCTCGTGGGATTAAGGTATCAGATTACAGCTATGGCAACCCAGCAAT 1726
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1727 TTAGTGGACAATAACATGAAATAATACCATATTTCTCGAGTAGGCGCAGCAATGGCG 1786
 QY 561 SerThrGluGluIleAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1787 TCCACAGAGGAAAAGCTTATGAGATCATGAGGAGCTCATGCTAGCTATGCTGTGTC 1846
 QY 581 IlePheGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600

DB 1847 ATTTTGGAGGCTCACTGGGTATTCCTCTGATGATCAACAAGTTCTTTGGATGGTC 1906
 QY 601 ArgIleGlyGlySerThrAspThrGlySerHisIleIleValLeuAsnAspTyrTyrThrPro 620
 DB 1907 CGGATTGGAGGAGCACAGATACAGGCAACATATCAGGAGAGTACTATTATATCTCCA 1966
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 DB 1967 ACTGGGAGTTCGCTGTCGACCGTGAAGTTCTCAGTCTGCTCAACTCCCTCATGTAC 2026
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaIysArgProGlyIlePhe 660
 DB 2027 AGATGTGTACTATCTGCTTTGGACAGGTTTACAGAGCCAGGCTCTCCAGGCTTT 2086
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 DB 2087 GACCGTGTCCGAATGCTGAGATTGGGAATAAGACTTTGAGCTTGATGTTGCTGGAGAA 2146
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValIysAspLeuAspAsnArg 700
 DB 2147 GGTATACACAGAACATTTGCTGCTCAGGATATACAGGTAAGGACCTGGATATATCGA 2206
 QY 701 GlyLeuSerArgThr 705
 DB 2207 GCGTTGTCAAGGACA 2221

RESULT 3
 BT007100 2118 bp mRNA linear PRI 13-MAY-2003
 LOCUS Homo sapiens integral membrane protein 1 mRNA, complete cds.
 DEFINITION
 ACCESSION BT007100
 VERSION BT007100.1 GI:30583038
 KEYWORDS FLI CDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2118)
 AUTHORS Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
 TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor vector
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2118)
 AUTHORS Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
 COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: http://bioinfo.clontech.com/orfclones.

FEATURES
 source 1..2118
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GH00707X1.0"
 /clone_lib="BD Creator(TM) CDS Library derived from MGC collection"
 /lab_host="DH5alpha Tl resistant"
 /note="Vector: pDNR-Dual"
 1..2118

CDS

/codon_start=1 /product="Integral membrane protein 1" /protein_id="AAP35764.1" /db_xref="GI:30583039" /translation="MTKFGFLRIGYERKQDILLKLLILSMAVLSTFSTRFAVLRPESV LHEPDPYNYRTTFLAEBEGYKPHNFDDRAWPLGRIIGGTIYPLGMITSAAIYHV LHPHTIDINVCVELAPLSPSSFTTIVHLKELKADAGALLAAALAVPGYKISR SVAGSDNEGIALFCMLTYVMKIKAVTGSICWAACALAYEVMYSSWGGYVPLINL IPLVLVLMLTRSHRIYVAYCTVYCLGILSLWQISFVGFQVLSERHMAAFVGL COIFAFVILSKNSPQFEVLFPSVLSLVGFVLLTVGALLMTGKLSIHWTRGFYSLL DPSYAKNNPIIASVSEHQPTWSSYFDLQILVFPFVGLYFCFNLSDARIFIMY GTSMYFAVMRLMLVLAVPMLISGIVSOVLSTYMKNLDSRPDKSKKQODSTY PIKNEAGSMLVMAFLITYFTSTVTSSEYSSPILSARGGGSGRIIFDDFREA YVHLRHTNPDAKVMWMDYGYQITAMANETILVNNNTWNTHSRVGQAMASTBKA YEIMRELDVSVLVTFGLGTGYSDDINKPLMVRIGGSTDTGKHIKENDYTPTEGF RVDREGSPVLLNCLMVKMYRFGQVTEAKRPPGPDVRVNAISGNKFDLVDLEAY TTEHVLVRIYKVLDDNRGLSRT"									
ORIGIN									
Alignment Scores:									
Pred. No.: 1.58e-226 Length: 2118									
Score: 226.00 Matches: 674									
Percent Similarity: 98.25% Conservative: 0									
Best Local Similarity: 98.25% Mismatches: 6									
Query Match: 32.06% Indels: 12									
DB: 9 Gaps: 0									
US-10-028-384-12 (1-705) x BT007100 (1-2118)									
QY	1	MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu	20						
DB	1	ATGACTAAGTTGGATTTTGGATTGTCCTATAGAGAGCAGGACACATTTTGAAGCTT	60						
QY	21	LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg	40						
DB	61	CTCATCTCTCAAGGGCTGCTATTATCTCTCCATCTGCTGCTGCTGCTGCTGCTGCTG	120						
QY	41	PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu	60						
DB	121	TTTGAAGTGTTATCCATGAGTTTGTATCCGTTACTTTAATATCCGACTACCAAGTTCTG	180						
QY	61	AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu	80						
DB	181	GCTGAGGAGGGGTTTATAAATCCATTAACCTGTTTGTATGACGAGCCTGTTACCTTTG	240						
QY	81	GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr	100						
DB	241	GGACGAATCATTTGGAGAGAACATTTACCAGGTTTATGATCATCCTCTGCTGCATCTAC	300						
QY	101	HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla	120						
DB	301	CATGTACTCCATTTTCCCATCCATCACCATCGACATTCGGAATGCTGTGTGTTCTTGCC	360						
QY	121	ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysA	140						
DB	361	CCTCTCTCTCTCTCTCTCCACAC-CATCGTCACGTACCA-CCTTACCAGAGCTCAAGG	418						
QY	140	spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA	160						
DB	419	ATGCAAGGGGCTGGGCTTCTGCTGTCATGATTAAGGGATTTGCACTTTTGCATGCTACCT	478						
QY	160	rgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr	180						
DB	479	GATCTGTGCTGCTCTCTATGATTAATGAAGGGATTTGCACTTTTGCATGCTACCT	538						
QY	180	YTYMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaL	200						
DB	539	ACTACATGTGGATCAAGGAGTAAGAGCTGGTTCATCTGTTGGGAGCTAAAGTGTGCC	598						
QY	200	euAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuIleP	220						
DB	599	TTGCTTTATTTCTACATGCTCTGTCATGGGAGGTTATGTTCTGCTGATCAACTTAAATC	658						
QY	220	roLeuHisValLeuValLeuMetLeuThrThrGlyArgPheSerHisArgIleTyrValAla	240						
DB	659	CTCTCCACGCTCTCTGCTGATGCTACAGGCGGTTTCTCTCACCGGATCTATGTGGCT	718						
QY	240	YrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe	259						
DB	719	ACTGTACTGTTTACTGCTGGGCACATATCTTCTAT-GCAGATCTCTCTTTGGGGTTTC	777						
QY	260	GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI	279						
DB	778	CAGCTGTCTCTTTCATCAGAGCACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAGAT	836						
QY	279	eHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPh	299						
DB	837	CCATGGCTTTGTGGATTACCTGGCGACAGATTGAATCCACAACAATTTGAAGTCTCTTT	896						
QY	299	eArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetIle	319						
DB	897	CCGAGAGGTCATCTCTCTGTGTAGGCTTTGTCTCTCACGTTGGAGCTCTCTCTATGCT	956						
QY	319	uThrGlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAl	339						
DB	957	GACAGGAAAAATATCTCTCTGGAGGGGGCTTTCTACTGCTCTGGATCCCTCTTATGC	1016						
QY	339	alyeAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSe	359						
DB	1017	TAAGAACACATCCCCATCATCTCTGTGTGTGAGCATCAGCCACCAACCTGGTCTCTC	1076						
QY	359	rTyrTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPh	379						
DB	1077	ATACTATTTTGACCTGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1136						
QY	379	eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPh	399						
DB	1137	TAGCAACCTGCTCTCATGCCGGATTTTATCATCATGATATGGTGTGACAGCATGACTT	1196						
QY	399	eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG	419						
DB	1197	TTGAGTGTAATGTTGGCTCTAATGCTAGTGTGGACCTGTTATGTG-CATTCTCTCTG	1255						
QY	419	lylleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProA	439						
DB	1256	GCATTTGAGTCTCCAGGCTGCTGTCCACATACATAGAGATCTGGACATAAGTCGTCAG	1315						
QY	439	spLysLysSerLysLysGlnGlnAspSerThrTyrProIleLysIle-GluValAlaSer	458						
DB	1316	ACAAAGAGAGCAAGAGCAAGGATTTCCACCTACCTACCTAATAAGAA-TGAAGTGGCAAGT	1374						
QY	459	GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpVal	478						
DB	1375	GGGATGATACCTGGCTGCT	1434						
QY	479	ThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer	498						
DB	1435	ACCAGTGAAGGCTACT	1494						
QY	499	ArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu	518						
DB	1495	AGGATCATATTATGATGATCTCCGAGAGCATATATTGGCTTCGTATATATCTCTCAGAG	1554						
QY	519	AspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg	538						
DB	1555	CATCGAAGGTCATGCT	1614						
QY	539	ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla	558						
DB	1615	ACAAATTTTAGTGGACAAATAACATGGAATAATACCATATTTCTCGAGTAGGCGAGCA	1674						
QY	559	MetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal	578						
DB	1675	ATGGCTGCCAGAGGAAAGGCTATGAGATCATGAGGAGGCTCGATGTACGTATGTG	1734						
QY	579	LeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrp	598						

Db 659 CTCTCCACGCTCTGCTGATGCTCACAGGCGGCTTCTCTCACCGGATCTATGTGGCCT 718
 Qy 240 yrcysThrValTyrCysLeuGlyThrIleLeuSerArg-glnIleSerPheValGlyPhe 259
 Db 719 ACTGACTGTTTACTGCTGGGCACTATACCTTTCTAT-GCAGATCTCTCTTGGGGTTTC 777
 Qy 260 GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuGlyGlnIle 279
 Db 778 CAGCCTGCTCTTCTATCAGACACATGCGACG-CTTTGGGGTCTTTGGTCTCTGCGAGAT 836
 Qy 279 ehislalpheValaspTyrLeuArgSerIleLeuAsnProGlnClnPheGluValLeuPh 299
 Db 837 CCATGCCCTTTGGGATTAACCTGCGGAGCAAGTTGAATCCCAACAAATTTGAAGTTCTTT 896
 Qy 299 eArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLe 319
 Db 897 CCGGAGCGTCACTCTCTGCTAGGCTTTGCTCTTCTCACCGTGGGAGCTCTCTCTATGCT 956
 Qy 319 uThrGlyIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAl 339
 Db 957 GACAGAAATATCTCTCTGGACGGGGCGTTTCTCTGCTGGATCCCTCTTATGC 1016
 Qy 339 alyAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSe 359
 Db 1017 TAAGAACAACATCCCACTATGCTCTGCTGATCAGCATCAGCCCAACCTGGTCTC 1076
 Qy 359 rTyrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPh 379
 Db 1077 ATACTATTTCACCTGCGAGCTCTCTGCTCTTCTATGCTTCCAGTGGCCTCTATTACTGCT 1136
 Qy 379 eSerAsnLeuSerAspAlaArgIlePheIleLeuMetTyrGlyValThrSerMetTyrPh 399
 Db 1137 TAGCAACCTGCTGATGCCGGATTTTATCATCATGTATGCTGACCAAGCATGTACTT 1196
 Qy 399 eSerAlaValMetAlaGluMetLeuValLeuAlaProValMetSer-IleLeuSerC 419
 Db 1197 TCACCTGATATGGTGGCTTAATGCTAGTGTGGACCTGTATGCTG-CATTCTCTCTG 1255
 Qy 419 lyleGlyValSerGlnValLeuSerThrTyrMetLeuAsnLeuAspIleSerArgProA 439
 Db 1256 GCATTGGAGTCTCCAGGCTGCTCCACATACATAGAGATCTGGACATAGTCTGCTCAG 1315
 Qy 439 spLyIleSerIleGlnGlnAspSerThrTyrProIleLeuIle-GluValAlaSer 458
 Db 1316 ACAGAGAGAGAGAGAGAGAGAGAGAGATCCACCTACCTATTAAGAA-TGAAGTGGCAAGT 1374
 Qy 459 GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpVal 478
 Db 1375 GGGATGATCTGCTCATGGCTTCTTCTCATCCTACCTACCTTCTCATCCTGGGTG 1434
 Qy 479 ThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer 498
 Db 1435 ACAGTGGAGGCTACTCTCTCCGTCATTTGCTATCTGCTGCTGGTGGGATGGCAGT 1494
 Qy 499 ArgIleIlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu 518
 Db 1495 AGGATCATATTGATGACTCTCCGAGAGCATATTATTGGCTTCTGCTATAATCTCCAGAG 1554
 Qy 519 AspAlaLeuValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg 538
 Db 1555 GATGCAAGGTCATCTCTGGTGGGATTTAGCTATCAGATCAGCTATGCGCAACCGA 1614
 Qy 539 ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla 558
 Db 1615 ACAATTTTATGGAGCAATAACACATAGGAATAATACCCATATTTCTCAGTAGGCGAGCA 1674
 Qy 559 MetAlaSerThrGluGlyLeuAlaTyrGlnIleMetArgGluLeuAspValSerTyrVal 578
 Db 1675 ATGGGCTCCAGAGAGAGAGAGAGAGAGATGATGATGATGAGGAGCTCGATGCTATGCT 1734
 Qy 579 LeuValIlePheGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTrp 598
 Db 1735 CTGGTCATTTTGGAGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1794

Qy 599 MetValargIleGlySerThrAspThrGlyIleHisIleIleIleGluAsnAspTyrTyr 618
 Db 1795 ATGTCGCGATGAGGAGACAGATACAGCAACATATCAAGGAGATGACTATAT 1854
 Qy 619 ThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeu 638
 Db 1855 ACTCAACTGGGAGTTCCGTGGACCGTGAAGGTTCTCCAGTGTGCTCACTGCTC 1914
 Qy 639 MetTyrIleMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaIleArgProPro 658
 Db 1915 ATGTACAGATGTGTTACTATCGCTTTGGACAGTTTACACAGACCCAGGCTCTCTCA 1974
 Qy 659 GlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeu 678
 Db 1975 GGCTTTGACCGTGTCCGAAATGCTGAGATTGGAAATAAGACTTTGAGCTTGATGCTCG 2034
 Qy 679 GluGlu 680
 Db 2035 GAGGAA 2040
 RESULT 5
 LOCUS CQ834708 2329 bp DNA linear PAT 29-JUL-2004
 DEFINITION Sequence 579 from Patent WO2004058805.
 ACCESSION CQ834708
 VERSION CQ834708.1 GI:50834245
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 MATSUDA, A. and YONETA, S.
 T cell activating gene
 TITLE
 JOURNAL Patent: WO 2004058805-A 579 15-JUL-2004;
 Asahi Kasei Pharma Corporation (JP)
 FEATURES
 source
 1. 2329
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 119..2080
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAH05569.1"
 /db_xref="GI:50834246"
 /translations="MTKFGFLRLSYEKQDTLLKILISMAAVLSFSTRDLFAVRFESV
 IHEPDPYNYRTTRFLAEGCFYKFNHFDRAWYPLGRIIGGTIVPGLMITSAAIYHV
 LHPHITIDRNVCVFLAFLFSFTTIVTYHLTKELKADAGLLAAATAVVPGVISR
 SVAGSDNEGIALFCMLTYMYIYVAYCTVYCTGLTSKQISFVGFQFVLSSEMAA
 FGPGLCQHAFVDYLRSLNPOQFVLRPSVISLVGFVLLITVGLLMLTGLISPTWG
 RPYLLDPSYAKNNIPIIASVSEHQPTWSSYIFDLQLLVFMPFVGLYFCFNLSDAR
 IFIMYGVTSNMFSAVVMRLMLVAPMCLISGIVSVLSTKXNLDLSRDPKSKK
 QDSTYPIKNEVASGMILYMAFFLITFTHTWTSVSEYSSPISLVARGGDSRIIF
 DDREAYWLHRTNPEDAKVMSWDYQITAMANTILVNNNTWNTHTISRVGAMA
 STEBEKIMRELDVSVLVI FGLTGYSSDDINKFLMWRIGGSTDTGKH KENDYY
 TPGEPVDRSEGSEVLNCLMYKVCYRFQGVTEAKRPPGDFRVRNAEIGNKDFELD
 VLEBAYTTEHWLVRIYKVDLDRGLSRI"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,71e-226 Length: 2329
 Score: 226.00 Matches: 407
 Percent Similarity: 98.55% Conservative: 0
 Best Local Similarity: 98.55% Mismatches: 3
 Query Match: 32.06% Indels: 6
 DB: Gaps: 0
 US-10-028-384-12 (1-705) x CQ834708 (1-2329)
 Qy 271 PheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeu 290

Db 773 TTGGGGTCTTTGGTCTCTGCCAGATCCATGCTTTTGGATTACCTGCGCAGAGTTG 832
 Qy 291 AsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPheValLeu 310
 Db 833 AATCCACACAAATTTGAAGTTCTTTCCGGAGGCTCATCTCTCTGGTAGGCTTTGCTCTT 892
 Qy 311 LeuThrValGlyAlaLeuLeuMetLeuThrGlyIleSerProThrPheArgPhe 330
 Db 893 CTCACCGTGGAGCTCTCTCATGCTGACAGGAAATATCTCCCTGGAGGGGGGTTTC 952
 Qy 331 TyrSer-LeuLeuAspProSerTyrAlaLysAsnAsnIleProIleIleAlaSerValSe 350
 Db 953 TACCC-GTGTGATCCCTCTTATGCTAAGACACACATCCCATCATTTGCTTCTGTGTC 1011
 Qy 350 rGluHisGlnProThrThrTrpSerSerTyrPheAspLeuGlnLeuValPheMe 370
 Db 1012 TGAGCATCAGCCCAACCTGGTCTCATACTATTTTGACCTGACGCTCCCTGCTTTCAT 1071
 Qy 370 tPheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleI 390
 Db 1072 GTTTCCAGTTGGCTCTTATTACTGCTTTACCACTGCTGATGCCGGAATTTTATCAT 1131
 Qy 390 eMetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuVal 410
 Db 1132 CATGTATGTTGACCAAGATGATCTTTTCCAGCTGAATGGTGGCTTCAATGCTAGTGT 1191
 Qy 410 uAlaProValMetSer-IleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrM 430
 Db 1192 GGCACCTGTTATG-G-CATTCTCTGCGCATTTGGAGTCTCCAGGTGCTGTCACATACA 1250
 Qy 430 etLysAsnLeuAspIleSerArgProAspLysIleSerLysGlnGlnAspSerThr 450
 Db 1251 TGAAGAATCGACATAGCTCGTCCAGACAGAGAGCAAGAGCAACAGGATTCACCT 1310
 Qy 450 yrProIleLysIle-GluValAlaSerGlyWeIleLeuValMetAlaPheLeuIle 469
 Db 1311 ACCCTATTAGAA-TGAAGTGGCAAGTGGATGATGATGCTGTCATGCTTTCTTCATC 1369
 Qy 470 ThrTyrThrPheHisSerThrTrpValThrSerGluAlaTyrSerSerProSerIleVal 489
 Db 1370 ACTTACACCTTTTCATTCAACTGGTGGTACAGTGGGCTTACTCTCTCCGTCCTATGTA 1429
 Qy 490 LeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyr 509
 Db 1430 CTATCTGCGCTGCTGGGATGGCAGTAGGATCATATTTGATGACTTCGAGAGACATAT 1489
 Qy 510 TyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrpAspTyrGly 529
 Db 1490 TATTGGCTTCGTATAATATCCAGAGGATGCCAAGGTCATGCTCTGGTGGGATTATGCC 1549
 Qy 530 TyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsn 549
 Db 1550 TATCAGATTACAGCTATGGCAACCCGAACTTTTGTAGTGGACATAACACATGGAATAAT 1609
 Qy 550 ThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluLysIleValTyrGluLe 569
 Db 1610 ACCCATATTTCTCAGTAGGGAGGCAATGGCTCCACAGAGAAAGCCATGAGATC 1669
 Qy 570 MetArgGluLeuAspValSerTyrValLeuValIlePheGlyIleLeuThrGlyTyrSer 589
 Db 1670 ATGAGGAGCTCGATGTCAGTATGCTGCTCATTTTGGAGGCTCATCGGTATTCC 1729
 Qy 590 SerAspIleAsnLysPheLeuTrpMetValArgIleGlySerThrAspThrGly 609
 Db 1730 TCTGATGATATCAACAAGTTCTTTTGGATGCTCCGATTGGAGGAGCAAGATACAGGC 1789
 Qy 610 LysHisIleLysGluAspTyrTyrThrProThrGlyGluPheArgValAspArgGlu 629
 Db 1790 AAACATATCAAGAGATGACTATTATCTCAACTGGGAGTTCGCTGGACCGTGAA 1849
 Qy 630 GlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGln 649
 Db 1850 GGTTCCTCAGTCTGCTCACTGCTCATGATACAGATGTTTACTATGCTTTGGACAG 1909

Qy 650 ValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaGluIleGly 669
 Db 1910 GTTTACACAGAGCCAGCGCTCTCCAGGCTTGACCGTGTCCGAAATGCTGAGATTGGG 1969
 Qy 670 AsnLysAspPheGluLeuAspValLeuGluIle 680
 Db 1970 AATAAGACTTTGAGCTTGATGTCCTCGAGGAA 2002
 RESULT 6
 CQ834706 2339 bp DNA linear PAT 29-JUL-2004
 LOCUS
 DEFINITION Sequence 577 from Patent WO2004058805.
 ACCESSION CQ834706
 VERSION CQ834706.1 GI:50834243
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 Matsuda, A. and Yoneta, S.
 T cell activating gene
 TITLE Patent: WO 2004058805-A 577 15-JUL-2004;
 JOURNAL Asahi Kasei Pharma Corporation (JP)
 FEATURES
 source 1..2339
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 29..595
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAH05568.1"
 /db_xref="GI:50834244"
 /translation="MLTYYMWIKAVKTSICWAAKALAYFYVWSWGQYVPLNLI
 PLHLVLMTRGRFSHRIVAVCTVYIGTILSQISFVGPQVYLSEHMAAFGVGLC
 OIHAFVDYLRSLKLPQDFVLFPSVLSLEKYLPRGVSTRCPWLRLTTPSLLLCL
 SISLPQGPHTILTCSSSSSCFQVQGLVF"
 CDS
 Alignment Scores:
 Pred. No.: 1,726-226 Length: 2339
 Score: 226.00 Matches: 306
 Percent Similarity: 98.71% Conservative: 0
 Best Local Similarity: 98.71% Mismatches: 2
 Query Match: 32.06% Indels: 4
 DS: 6 Gaps: 0
 US-10-028-384-12 (1-705) x CQ834706 (1-2339)
 Qy 373 ValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleMetTyr 392
 Db 765 GTTGGCTCTATTACTGCTTTAGCAACCTGCTCATGCCCGGATTTTATCATCATGTAT 824
 Qy 393 GlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeuAlaPro 412
 Db 825 GGTGTGACCAAGCATGATCTTTTACGTGTAAATGCTGCTTAATGCTAGTGTGGACCT 884
 Qy 413 ValMetSer-IleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAs 432
 Db 885 GTTATGTG-CATTCTCTGCGATTGGAGTCTCCAGGTGCTGCCACATACATGAAGA 943
 Qy 432 nleuaspIleSerArgProAspLysLysSerLysGlnAspSerThrTyrProI 452
 Db 944 TCTGGACATAAGTCTGTCAGACAGAGCAAGAGCAAGGATTCACCTACCTAT 1003
 Qy 452 elysIle-GluValAlaSerGlyMetIleLeuValMetAlaPheLeuIleThrTyr 472
 Db 1004 TAAGAA-TCAAGTGGCAAGTGGATGATGATGCTGCTGCTTCTTCTCATCACCTACA 1062
 Qy 472 hrPheHisSerThrTrpValThrSerGluAlaTyrSerSerProSerIleValLeuSerA 492

```

Db      1063  CTTTTCAATCAACCTGGGTGACAGTGGAGGCTACTCTTCTCCGTCCATTGTACTATCTG 1122
Qy      492  laargglycllyaspGlySerArglllellePheAspAspPheArgGluAlaTyrrTrrPl 512
Db      1123  CCGGTGGTGGGATGGCAGTAGGATCAATTATGTAGTCTCCGAGAGCATAATTATGGC 1182
Qy      512  euArgHisAsnThrProGluAspAlaLysValMetSerTrrTrrPAspTyrrGlyTyrrGlnI 532
Db      1183  TTCGTCAATAACTCCAGAGGATGCGAAGGTCATGTCGTGGGATTATGGCTATCAGA 1242
Qy      532  leThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrrPAsnAsnThrHisI 552
Db      1243  TTACAGCTATGCAACCAACCAACAATTTAGTGGCAATAACACATGGAAATAATACCCATA 1302
Qy      552  leSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaLysValGluIleMetArgG 572
Db      1303  TTTCTCGAGTAGGAGCAATGGGCTCCACAGAGGAAAAGCCTATGAGATCATGAGGG 1362
Qy      572  luLeuAspValSerTyrrValLeuValIlePheGlyGlyLeuThrGlyTyrrSerSerAspA 592
Db      1363  ACTCGATGTCAGCTATGTCGTGTCATTTTGGAGGCTCAGTGGGTATTCCTCTGATG 1422
Qy      592  spIleAsnLysPheLeuTrrMetValArgIleGlyGlySerThrAspThrGlyLysHisI 612
Db      1423  ATATCAACAAGTTCTTTGGATGGTCCGATTTGGAGGAGCACACATACAGGCAACATA 1482
Qy      612  leLysGluAsnAspTyrrTyrrThrProThrGlyGluPheArgValAspArgGluGlySerP 632
Db      1483  TCAGAGAGNAGTATTAATCTCCAACTGGGAGTTCCGTGGAGCCGTGAAGGTTC 1542
Qy      632  roValLeuLeuAsnCysLeuMetTyrrLysMetCysTyrrTyrrArgPheGlyGlnValTyrr 652
Db      1543  CAGTGTGCTCAACTGCCTCATGTACAAAGATGTGTACTATCGCTTTGGACAGGTTTACA 1602
Qy      652  hrGluAlaLysArgProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysA 672
Db      1603  CAGAAGCCAGGCTCTCCAGCTTGACCGGTGCCGAAATGCTGAGATGGGAATAAG 1662
Qy      672  spPheGluLeuAspValLeuGluGlu 680
Db      1663  ACTTTGAGCTTGATGTCCTCGAGGAA 1688

RESULT 7
LOCUS   CQ834710
DEFINITION Sequence 581 from Patent WO2004058805.
ACCESSION CQ834710
VERSION   CQ834710.1
KEYWORDS Homo sapiens (human)
SOURCE   Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Matsuda, A. and Yoneta, S.
TITLE T cell activating gene
JOURNAL Patent: WO 2004058805-A 581 15-JUL-2004;
Asahi Kasei Pharma Corporation (JP)
FEATURES
Source
1..2435
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
73..2187
/note="unnamed protein product"
/codon_start=1
/protein_id="CAH05570.1"
/db_xref="GI:50834248"
/translation="MTKFGRLSYEQDILLKLLLSVAALVSFTRFLFAVLRRESV
IHEPDFNYETTRFLAEFGYKPHNFDRAWYPLRIGGTIYPLMTSAALYHV
LHFRITIDNVCVFLAPFSSTITVYHLTKELKADAGALLAAVAVPGVISR
SVAGSYDNEGIAIFCMLLYTMMIKAVKTSICWAAKCALAFYWSWGGYVFLNLI
PLRVLMMLTRFSHRIVAYCTVCLGTLISMQISFVGFPVLSSEHMAAGVFLCLC

```

CDS

```

OIHAFVDYLBKLNLPQCFEVLPRSVLSLVGFVLLTYGALLMLTGKISPMWGRFYSLLD
PSAKNNPIIASVSEHPTTSSSYFDLQLVFMEFVGLTYCFNSLSARLPIIMYG
VTSNYFAVMRLMLVLAPVMGLSGVSLSTYKKNLDSRPDKSKKQDDSTYP
IKNEVASGMLVAFELITPTSTWTSEAVSSPSIVLSARGGDSRIIFDDPREAY
YMLRNPEDAKVMWDYQITAMANRTILVDNNTWNTTHISRVGQAMASTEKAY
EIMRLDVSYLVIFFGLTGYSDDKLFWRIIGSTDTGKHIKENDYPTGFEFR
VDREGSPVLNCLMYKCYVRFQVTEAKRPPGDFVRNABIGNKDFELDLVEAYT
TEHWLVRIYKVKLDNRGLSRT"
ORIGIN
Alignment Scores:
Pred. No.: 1.78e-226 Length: 2435
Score: 226.00 Matches: 671
Percent Similarity: 97.67% Conservatives: 0
Best Local Similarity: 97.67% Mismatches: 9
Query Match: 32.06% Indels: 16
DB: 6 Gaps: 0
US-10-028-384-12 (1-705) x CQ834710 (1-2435)
Qy      1  MetThrLysPheGlyPheLeuArgLeuSerTyrrGluLysGlnAspThrLeuLeuLysLeu 20
Db      73  ATGACTAAGTTTGGATTTTTCGATTTGCTCTATGAGAGCAGGACACACTTTTGAAGCTT 132
Qy      21  LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db      133  CTCATTCTGTCAATGGCTGTGTATTATCTCTCCACTCGTCTGTCTGCTGCTGAGA 192
Qy      41  PheGluSerValIleHisGluPheAspProTyrrPheAsnTyrrArgThrThrArgPheLeu 60
Db      193  TTTGAAAGTGTATTCATGAGTTTGTATGCTGCTCTTTAATATCGGACTACCAAGTTTCTG 252
Qy      61  AlaGluGluGlyPheTyrrLysPheHisAsnTrrPheAspAspArgAlaTrrTyrrProLeu 80
Db      253  GCTGAGAGGGGTTTATAAATTCATGCTGTTGTGACCGAGCCCTGTTACCTTTG 312
Qy      81  GlyArgIleleGlyGlyThrIleTyrrProGlyLeuMetIleThrSerAlaAlaIleTyrr 100
Db      313  GGAGCAATCATTTGGAGGAACAATTTACCCAGGTTTAATGATACCTCTCTGCTCAATCTAC 372
Qy      101  HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db      373  CATGTACTCCATTTTTCACATCACATCGACATTCGGAAATGCTGTGTGCTGCTGCCC 432
Qy      121  ProLeuPheSerSerPheThrSerIleValThrTyrrLeuLeuThrLysGluLeuLysA 140
Db      433  CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 490
Qy      140  spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrrIleSerA 160
Db      491  ATGACAGGGGCTGGGCTTCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550
Qy      160  rgSerValAlaGlySerTyrrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 180
Db      551  GATCTGTGGCTGGCTCTCTATGATAAAGAGGATTCCTCTTTGTCATGCTACTCACCT 610
Qy      180  YrTyrrMetTrrIleLysAlaValLysThrGlySerIleCysTrrPAlaAlaLysCysAla 200
Db      611  ACTACATGTGATCAAGGAGTAAAGACTGGTTCATCTGTTGGGAGCTAAGTGTGCCC 670
Qy      200  euAlaTyrrPheTyrrMetValSerSerTrrpGlyTyrrValPheLeuIleAsnLeuLeu 220
Db      671  TTG--CT--TTCTACATGGTCTCGTCATGGGAGGTTATGTGTTCTCTGATCAACTTAATTC 727
Qy      220  roLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrrValAla 240
Db      728  CTCTCCAGCTCTCTGTGCTGATGCTCACAGGCCGTTTCTCTCACCGATCTATGTGGCT 787
Qy      240  YrCysThrValTyrrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPhe 259
Db      788  ACTGTACTGTATTACTCTGGGACATATCTTCTAT--GCAGATCTCTCTTTTGTGGTTC 846
Qy      260  GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279

```

Db 847 CAGCGTCTCTTTCATCAGACCATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAGAT 905
 Qy ehissalaphelValaspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuph 299
 Db CCATGCCCTTTGGGATTAACCTGCGCAGCAAGTTGAATCCACAAATTTGAAGTCTCTTT 965
 Qy eArgSerValLleSerLeuValGlyPheValLeuThrValGlyAlaLeuLeuMetle 319
 Db CCGAGAGCGTCATCTCTCTGTTAGGCTTTGCTCTTCTCACCGTGGGAGCTCTCTCATGCT 1025
 Qy uThrGlyLysLleSerProThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAl 339
 Db GACAGAAATAATCTCCCTGGAGCGGGCGTTTCTACTCGCTCGTGGATCCCTCTTATGTC 1085
 Qy alyAsnAsnLleProLleLleAlaSerValSerGluHisGlnProThrThrTrpSerSe 359
 Db TAAGAACACATCCCATCATCTGCTTCTGTGCTGTGAGCATCAGCCACACCTGGTCTCTC 1145
 Qy rTyrTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTy-Cyaph 379
 Db ATACTATTTTGAACCTGCGAGCTCTCTGCTTCTGCTTCTGCTTCTGCTTCTTATGCTT 1205
 Qy eSerAsnLeuSerAspAlaArgLlePheLleLleMetTyrGlyValThrSerMetTyrPh 399
 Db TAGCAACCTGTCTGATGCCCGGATTTTATCATCATGTATGTTGGTGGACCTTATGTG-CAATCTCTCTG 1265
 Qy eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG 419
 Db TTCACCTGTATGTGGTGTCTAATGCTAGTGTGGACCTTATGTG-CAATCTCTCTG 1324
 Qy lylleGlyValSerGlnValLeuSerThrTyrMet-LysAsnLeuAspLleSerArgPro 438
 Db GCATTGGAGTCTCCAGTGTGCTGCTCCACATACAA-GAAGAATCTGACATAAGTCTGCTCA 1383
 Qy AspLysLysSerLysLysGlnInAspSerThrTyrProLysLysLysLysLysLysLys 458
 Db GACAAAGAGACAGACAGACAGATCCACCTACCTACCTATTAAGAA-TGAAGTGGCAAG 1442
 Qy rGlyMetLleLeuValMetAlaPhePheLleLleThrTyrThrPheHisSerThrTrpVa 478
 Db TGGGATGATATGTCATGCTGCTTCTTCTCATCATCATCATCATCATCATCATCATCAT 1502
 Qy lThrSerGluAlaTyrSerSerProSerLleValLeuSerAlaArgGlyLysAspGlySe 498
 Db GACCAGTGGGCTACTCTCTCGTCCATGTTACTATCTGCCCGTGGTGGGATGGCAG 1562
 Qy rArgLleLlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGl 518
 Db TAGGATCATATTGATGACTTCCGAGAGCATATTATGCTTCTGTCATTAATACTCCAGA 1622
 Qy uAspAlaLysValMetSerTrpTyrAspTyrGlyTyrGlnLleThrAlaMetAlaAsAr 538
 Db CGATCGGAAGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1682
 Qy rThrLleLeuValAspAsnAsnThrTrpAsnAsnThrHisLleSerArgValGlyGlnAl 558
 Db ACAATTTTATGTCACATACATGGAATATATACCATATTTCTCGAGTAGGCGAGGC 1742
 Qy aMetAlaSerThrGluLysAlaTyrGluLleMetArgGluLeuAspValSerTyrVa 578
 Db AATGCGTCCACAGAGGAAAGGCTATGAGATCATGAGGAGCTCGATGTCAGCTATGT 1802
 Qy lLeuValLlePheGlyLysLeuThrGlyTyrSerSerAspAspLleAsnLysPheLeuTr 598
 Db CTGCTGCTATTTTGGAGGCTCTACTGGTATTTCTCTGATGATATCAACAGTTCTTTG 1862
 Qy pMetValArgLleGlySerThrAspThrGlyLysHisLleLysGluAspAspTyrTy 618
 Db GATGTCGCGATTGGAGGAGCAGATACAGGCAACATATCAAGAGAAATGACTATTA 1922
 Qy rThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCySe 638

Db 1923 TACTCCACTGGGGAGTTCCGTGTGGACCGTGAAGTTTCTCCAGTGTCTCAACTGCCT 1982
 Qy uMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgPro 658
 Db CATGTACAGATGTGTCTACTATCGCTTTGGACAGTTTACACAGACCGAGCTCTCC 2042
 Qy oGlyPheAspArgValArgAsnAlaGluLleGlyAsnLysAspPheGluLeuAspVal 678
 Db AGCCTTTGACCTGTCGGAATGCTGAGATTGGAATAGAGACTTTGAGTTGATGCTCT 2102
 Qy uGluGlu 680
 Db GGAGGAA 2109
 RESULT 8
 BC048348
 LOCUS
 DEFINITION Homo sapiens integral membrane protein 1, mRNA (cDNA clone IMAGE:5561615), partial cds.
 ACCESSION BC048348
 VERSION BC048348.1
 KEYWORDS GI:29477243
 SOURCE Homo sapiens (human)
 ORGANISM
 REFERENCE
 1 (bases 1 to 2450)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettman, M., Madan, A., Young, A.C., Shrivastava, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shrivastava, S., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywicki, M.I., Skalska, U., Smal, U., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 2450)
 Strausberg, R.
 Direct Submission
 Submitted (07-MAR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Ackner, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantropop, S., Thomas, P.J., Touchman, J.W.,

Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 93 Row: p Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 22749414.
Location/Qualifiers

FEATURES

1..2450
/organism="Homo sapiens"
/no._type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5561615"
/tissue_type="Duodenum, adenocarcinoma"
/clone_lib="NIH MGC_88"
/lab_host="DH10B"
/note="vector: PCMV-SPORT6"
1..2450
/gene="ITMI"

gene

/note="synonyms: TMC, MGC9042"
/db_xref="LocusID:3703"
/db_xref="MIM:601134"
1..2163
/gene="ITMI"

CDS

/codon_start=1
/product="ITMI protein"
/protein_id="AAH48348.1"
/db_xref="GI:29477244"
/db_xref="LocusID:3703"
/db_xref="MIM:601134"
/translation="SPADRRLPPIHVMTKGFRLSVEKQDTLLKLLISMAAVAL
SFSTRUPAVLPSPFSDYFNRTTRELAEFGYKHNWDFDRAYPLGRIGGT
IYFLMITSAAIHVLHFPHITDIRNVCFAPLPSFTTIVTYHLTKELKADAGL
LAAMTAVVPGYSRSVAGSDNEGAI FCMILTYTMMIKAVKTSICWAAKCALAYF
YMYSSWGVYFLNLIPLHVLMLTGRFSHRYVACTVYCLGTLISMQISFVGQF
VLSEHMAAGFVGLCOIHAFVDYLRSLNPOQFVLSRVSISLGVLLTVGALLML
TGKISPTWGRFYSILQPSYAKNIPILIASUSEHPTWSSYFDLQLLVFMPFVGLY
CFNLSLDRPIELIMYGVTSMYFSAMVRLMLVLPVCLISGIGVSQVLSITMKNLDI
SRDPSKQODSTPIKNEVASGMILWNAFFLITYFTSHTSAYSPISVILSAR
GGGSRIFIDFEAYIYWLNRHNPEDAKVMSWDYGYQITAMANTRLVNDNTWNTH
ISRVQMASTPEKAYIEMRELDVSLVLI FGLGTGYSDDINDKFLMYRIIGSDTIG
KHKENDYYPTGEFRVREGSPVLLNCLMYKCYVRFGQVTEAKRPGFDRVNAE
IGNKDFELDLVEAYTTEHLVRIYKVKDLNRLSRT"

ORIGIN

Alignment Scores:

Pred. No.: 1,79e-226 Length: 2450
Score: 226.00 Matches: 674
Percent Similarity: 98.25% Conservative: 0
Best Local Similarity: 98.25% Mismatches: 6
Query Match: 32.06% Indels: 12
DB: Gaps: 0

US-10-028-384-12 (1-705) x BC048348 (1-2450)

Qy 1 MetThrlsPheGlyPheLeuArgLeuSerTyrGluYsGlnAspThrLeuLeuLysLeu 20
Db 46 ATGACTAAGTTTGGATTTTGGATTGCTCTATGAGACGACGACACATTTTGAAGCTT 105
Qy 21 LeuLeuSerMetAlaValLeuLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 106 CTCACTCTGCAATGCTGCTGATTTATCTCTCCACTCGTCTGTTGCTGCTCTGAGA 165
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 166 TTGTGAAGTGTATCAATGATTTGATTCGTACTTTTAATATCGGACTACAGGTTCTCTG 225
Qy 61 AlaGluGlyPheTyrIlePheHisAsnTrpPheAspArgAlaTrpTyrProLeu 80
Db 226 GCTGAGGAGGGGTTTATAAATCCATAACTGTTTGTATGACCGAGCGCTGGTACCTTTG 285
Qy 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100

Db 286 GAGCAATCATTTGGAGAAACAATTTACCCAGGTTTAAATGATCACCTCTCTGCAATCTAC 345
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 346 CATGTACTCCATTTTTCACATCCACATCGCATTCGGAATGTCTGTGTGTCTCTGGCC 405
Qy 121 ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrIleGluLeuLysA 140
Db 406 CCT 463
Qy 140 SPAlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
Db 464 ATGCAGGGGCTGGGCTTCTTGTCTGCCATGATGCTGTAGTCTCTGTGATATATCTCC 523
Qy 160 rGserValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrT 180
Db 524 GATCTGTGCTGCT 583
Qy 180 YrTyrMetTrpIleLysAlaValIleValIleValIleValIleValIleValIleVal 200
Db 584 ACTACATGTGGATCAAGGCGATTAAGACTGGTTCTCATCTGTGGCAGCTAAGTGTGCC 643
Qy 200 euAlaTyrPheTyrMetValSerSerTrpGlyTyrValPheLeuIleAsnLeuIleP 220
Db 644 TTGCTTATTCTTACATGGTCTCTCTCATGGGAGGTTATGTGTCTCTGATCACTTAATTC 703
Qy 220 rGluHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaI 240
Db 704 CTCTCCAGCTCTCTGT 763
Qy 240 YrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
Db 764 ACTGTACTGTTTACTGCTGGGCACTATACITTTCTAT-GCAGATCTCTTTTGTGGTTTC 822
Qy 260 GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279
Db 823 CAGCTGTCTCTTTCATCAGACACATGGCAGC-CTTTGGGGTCTTTGTCTCTCTCCAGAT 881
Qy 279 eHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuP 299
Db 882 CCATGCTTTTGTGATTTACCTCGGAGCAAGTTGAATCCACAATTTGAAGTTCCTTTT 941
Qy 299 eArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLe 319
Db 942 CCGGAGCGTCATCTCTGTGTAGGCTTGTCTCTCACCGTGGAGCTCTCTCTCATGCT 1001
Qy 319 uThrGlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAl 339
Db 1002 GACAGAAAAATATCTCTCTGGAGCGGGCGCTTCTACTCGCTGTGTGATCCCTCTTATGC 1061
Qy 339 aLysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSe 359
Db 1062 TAAGAACAAATCCCATCATTTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1121
Qy 359 rTyrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPh 379
Db 1122 ATACTTTTTCAGCTGAGCTCTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1181
Qy 379 eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPh 399
Db 1182 TAGCAACCTGCTCATGCGCGGATTTTATCATCATGTATGGTGTGACGACGATGACTT 1241
Qy 399 eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG 419
Db 1242 TTCAGCTGTAAATGGTGGCTTAATGCTAGTGTGGCACCTGTATGTG-CATTCTCTCTG 1300
Qy 419 lYlleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProA 439
Db 1301 GCATTGAGTCTCCAGAGTCTGTCCACATACATGAAATCTGGACATGAAGTCTGCTCCAG 1360
Qy 439 sPLysLysSerLysLysGlnGlnAspSerThrTyrProIleLysIle-GluValAlaSer 458

```

Db 1361 ACAAGAAGAGCAGCAAGCAAGGATCCACCTACCTATTAAAG-TGAAGTGGCAAGT 1419
Qy 459 GlyMetIleLeuValMetAlaPhePheLeuIleThrTyThrPheHisSerThrTrpVal 478
Db 1420 GGGATGATACCTGGTCATGGCTTTCTTCTCATCCTACCTACCTTTTCATCACTGGGGT 1479
Qy 479 ThrSerGluAlaIleTySerSerProSerIleValLeuSerAlaIleGlyGlyAspGlySer 498
Db 1480 ACCAGTGGGGCTACTCTCTCCGTCCTACTATGTTACTATCTGCTGCTGGTGGGATGGCAGT 1539
Qy 499 ArgIleIlePheAspAspPheArgGluAlaTyThrTrpLeuArgHisAsnThrProGlu 518
Db 1540 AGGATCATATTGATGACTCCGAGAGCAATATTATGGCTTCGCTATATCTACCTCAGAG 1599
Qy 519 AspAlaIleValMetSerTrpTrpAspTyGlyTyGlnIleThrAlaMetAlaAsnArg 538
Db 1600 CATCGAAGGTCATGCTCTGCTGGGATATTGGCTATCAGATTACAGCTATGGCAACCGA 1659
Qy 539 ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla 558
Db 1660 ACAATTTAGTGGCAATATACACATGGATATATACCATATTTCTCGATAGGCAAGCA 1719
Qy 559 MetAlaSerThrGluGluLysAlaIleTyGluIleMetArgGluLeuAspValSerTyVal 578
Db 1720 ATGGCGTCCACAGAGGAAAGCCATATCAGATCATGAGGAGCTCGATGCTCAGCTATGTG 1779
Qy 579 LeuValIlePheGlyGlyLeuThrGlyTyThrSerSerAspAspIleAsnIleTyPheLeuTrp 598
Db 1780 CTGGTCATTTTGGAGGCTTCATGGGTATTCCTCTGATGATATCAACAGTTCTTTGG 1839
Qy 599 MetValArgIleGlyGlySerThrAspThrGlyGlyHisIleIleGlyGluAsnAspTyTy 618
Db 1840 ATGGTCCGATTGGAGGAGCACAGATACAGGCAAAATATCAAGGAGATCACTATTAT 1899
Qy 619 ThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeu 638
Db 1900 ACTCCACTGGGAGTTCGGTGGACCGTGAAGGTTCTCCAGTGTCTCACTGCTC 1959
Qy 639 MetTyIleMetCysTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 658
Db 1960 ATGTACAGATGTCTACTATCGTTTGGACAGGTTTACACAGAACCAAGGCTCCTCCA 2019
Qy 659 GlyPheAspArgValArgAsnAlaGluIleGlyAsnIleAsnIleAsnIleAsnIleAsn 678
Db 2020 GGCTTGACCTGTCGCAATGCTGAGATTGGGATTAAGACTTTGAGCTTGATGCTG 2079
Qy 679 GluGlu 680
Db 2080 GAGGAA 2085

RESULT 9
CQ834714 2458 bp DNA linear PAT 29-JUL-2004
DEFINITION Sequence 585 from Patent WO2004058805.
ACCESSION CQ834714
VERSION CQ834714.1 GI:50834251
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Matsuda, A. and Yoneta, S.
TITLE T cell activating gene
JOURNAL Patent: WO 2004058805-A 585 15-JUL-2004;
Asahi Kasei Pharma Corporation (JP)
FEATURES
source
1..2458
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
93..2210
/note="unnamed protein product"
CDS

```

ORIGIN

```

Alignment Scores: 1.8e-226 Length: 2458
Pred. No.: 226.00 Matches: 674
Percent Similarity: 98.25% Conservative: 0
Best Local Similarity: 98.25% Mismatches: 6
Query Match: 32.06% Indels: 12
DB: 6 Gaps: 0

US-10-028-384-12 (1-705) x CQ834714 (1-2458)
Qy 1 MetThrIlePheGlyPheLeuArgLeuSerTyGluLysGlnAspThrLeuLeuLysLeu 20
Db 93 ATGACTAAGTTTGGATTTTGGATTGTCCTATGAGAGCAGGACACACTTTTGAAGCTT 152
Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 153 CTCATTCTGCTCAATGGCTGCTGTTATTCCTTCTCCACTGCTGCTGCTGCTGCTG 212
Qy 41 PheGluSerValIleHisGluPheAspProTyPheAsnTyArgThrArgPheLeu 60
Db 213 TTGGAAGTGTATCCATGAGTTTGCATCGTACTTAAATATCGGACTTACAGGTTCTCT 272
Qy 61 AlaGluGluGlyPheTyIlePheHisAsnTrpPheAspAspArgAlaValTyTyProLeu 80
Db 273 GCTGAGAGGGGTTTATAATTCATTAATTCATTAATTCATTAATTCATTAATTCAT 332
Qy 81 GlyArgIleIleGlyGlyThrIleTyProGlyLeuMetIleThrSerAlaAlaIleTy 100
Db 333 GGACGAATCAITGGAGGAACTTTTACCAGGTTTAAATGATCATCCTCTGCTGCAATCTAC 392
Qy 101 HisValIleHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 393 CATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGCTGTGTGTTCTGCTGCC 452
Qy 121 ProLeuPheSerSerPheThrSer-IleValThrTyLeu-LeuThrIleGluLeuLysA 140
Db 453 CCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 510
Qy 140 spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyTyIleSerA 160
Db 511 ATGACGGGCTGGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 570
Qy 160 rGSerValIleGlySerTyAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 180
Db 571 GATCTGTGCTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 630
Qy 180 TyTyMetTrpIleLysAlaValIleTyThrGlySerIleCysTyTrpAlaAlaLysCysAla 200
Db 631 ACTACATGTGGATCAAGGCAAGTAAAGACTGGTTTCCATCTGTTGGGAGCTAAGTGTGCC 690
Qy 200 eunlatTyPheTyMetValSerSerTrpGlyTyValPheLeuLeuLeuLeuLeuLeu 220
Db 691 TTGCTTATTTCATCATGTTCTGTCATGGGAGGTTATGTGTTTCTGATCAACTTAATTC 750
Qy 220 rOleuHisValLeuValLeuMetLeuThrGlyTyArgPheSerHisArgIleTyValAla 240

```

```

/codon_start=1
/protein_id="CAH055572.1"
/db_xref="GI:50834252"
/translation="MTKEGFRLSYEVKQDTLLKLLLSMAAVLSFSRLEFAVLPFESV
LHEFFITDIRNVCFVFLAPLFSSFTTVYHLTKELDKADAGLLAAAMIAVPGYISR
SVAGSYDNEGTAIFCMLLTYWMTKAVGTSGICWAACKALAYFYWSSWGYYFLINL
IFLHVLMLTGRFSHRIYVAYCTVCLGTILSMQISFVGFPQVLSSEHMAAFGVFLG
CCHAFVDYLRSLNPOQFELFRSVISLVGLVLLTVGALLMLTKISPTWGTGFYSLL
DSYAKNNIPITASVSEHPTWSSYVFDLOLVFMPVGLYCYFSLSDARIFIMY
GYTSVFAVWVRMLVLAIPWMLISGIGVSQVLSYMKLIDSRDPKSKKQODSTY
PIKNSVASCMLLWAPFLITYTHSTWTSYSSPSLVLSARGGDSRILIPDPREA
IYWLKRNHPEDAKXNMWDIGYQITANRILVDNNTWNTHTHSRVQMAVSTESKA
YIMELDVSVLVLFVFGGLTYSSDDINKFLMWVRIGGSDTDGKHKEKENDYYIPTGF
RYDRGSPVYVLCNLYKMYCYRFGVYVTEAKRPFGRVNRNAEIGNKDFELDLVEAY
TTEHWLVRIYVKDLNDRGLSRT"

```

Db 751 CTCTCCAGCTCTCTGCTGATGCTCACAGCGGTTTCTCTCACCGGATCTATGCGCT 810
Qy 240 YrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
Db 811 ACTGTACTGTTTACTGCTGGGCACTATCTCTCTAT-GCAGATCTCTCTTGTGGGTTTC 869
Qy 260 GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnIle 279
Db 870 CAGCGTGTCTTTCATCAGACACATGCGCAGC-CTTTGGGGTCTTTGGTCTCTGCGCAGAT 928
Qy 279 eHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPh 299
Db 929 CCATGCGCTTTGTGATTAACCTGCGCAGCAAGTTGAATCCACAAATTTGAAGTCTTTT 988
Qy 299 eArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLe 319
Db 989 CCGAGCGCTCATCTCTCTGTAGGCTTTGCTCTCTCACGCGGAGCTCTCTCATGCT 1048
Qy 319 uThrGlyLysLeuSerProThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAl 339
Db 1049 GACAGGAAATATCTCCCTGGAGCGGGCGTTTCTACTCGCTGCTGGATCCCTCTTATGC 1108
Qy 339 aLysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSe 359
Db 1109 TAAGAACACATCCCATCATCTGCTGTCTGCTGACATCAGCCCAACACCTGGTCCCTC 1168
Qy 359 rTyrTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPh 379
Db 1169 ATACTATTTGACCTGCGAGCTCTCTGCTCTCATGTTTCCAGTTGGCTCTTACTGCTT 1228
Qy 379 eSerAsnLeuSerAspAlaGlyPheIleIleMetTyrGlyValThrSerMetTyrPh 399
Db 1229 TAGCAACCTGCTGATCCCGGATTTTATCATCATGATGATGGTGACCAAGCATGACTT 1288
Qy 399 eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG 419
Db 1289 TTCAGCTGTAATGTGGCTCTAATGCTAGTGTGGCACCCTGTTATGTG-CATTCTCTCTG 1347
Qy 419 lYleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProA 439
Db 1348 GCATTGAGTCTCCAGGCTCTGTCACATACATGAAATCTGGACATAAGTCGTCACAG 1407
Qy 439 sPlyLysSerLysGlnAspSerThrTyrProIleLysIle-GluValAlaSer 458
Db 1408 ACAAGAGACAAAGACACAGATTCACCTACCTACCTACCTTTCATTAAGAA-TGAAGTGGCACT 1466
Qy 459 GlyMetIleLeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTrpVal 478
Db 1467 GGGATGATCTGCTCATGGCTTTCTTCTCATCACCTACCTTTCATTAACCTGGGTG 1526
Qy 479 ThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer 498
Db 1527 ACCAGTGAGCCCTACTCTCTCCGTCATTTGATCTATCTGCGCGTGGTGGGATGCGAGT 1586
Qy 499 ArgIleIlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu 518
Db 1587 AGGATCATATTTGATGACTCCGAGAGCATATATTGGCTCGTCATATACTCCAGAG 1646
Qy 519 AspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaSerArg 538
Db 1647 GATCGAAGTCTATGCTCTGGTGGGATTTATGGCTATCAGATTACAGCTATGCAACACGA 1706
Qy 539 ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla 558
Db 1707 ACAATTTAGTGACAAATACACATGAAATATACCATATTTCTCGAGTAGGGAGGCA 1766
Qy 559 MetAlaSerThrGluGlnLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal 578
Db 1767 ATGGCTCCACAGAGAAAGCCCTATGAGATCATGAGGAGCTGATGCTCAGCTATGCTG 1826
Qy 579 LeuValIlePheGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTrp 598

Db 1827 CTGGTCATTTTGGAGGCTCTACTGGGTATTCCTCTGATGATATCAACAAGTTTCTTTGG 1886
Qy 599 MetValArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyr 618
Db 1887 ATGGTCCGATTTGGAGGAGCAGATACAGCAACATATCAAGGAGATCACTATTAT 1946
Qy 619 ThrProThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeu 638
Db 1947 ACTCCAACTGGGAGTTCCGTGTGACCGTGAAGTTTCCAGTCTGCTCAACTGGCTC 2006
Qy 639 MetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProPro 658
Db 2007 ATGTACAGATGTTTACTATCTGCTTTGGACAGGTTTACACAGAGCCAGGCTCTCCA 2066
Qy 659 GlyPheAspArgValArgAsnNlaGluIleGlyAsnLysAspPheGluLeuAspValLeu 678
Db 2067 GGCCTTACCGGTGTCGGAATCTGAGATTGGGAATAAGACTTTGAGCTTGATGCTGCT 2126
Qy 679 GluGlu 680
Db 2127 GAGGAA 2132
RESULT 10
LOCUS CQ727867 2472 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 13801 from Patent WO02068579.
ACCESSION CQ727867
VERSION CQ727867.1 GI:42295093
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 13801 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
1..2472
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 1,81e-226 Length: 2472
Score: 226.00 Matches: 673
Percent Similarity: 98.25% Conservative: 0
Best Local Similarity: 98.25% Mismatches: 6
Query Match: 32.06% Indels: 12
DB: Gaps: 0
US-10-028-384-12 (1-705) x CQ727867 (1-2472)
Qy 2 ThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeuLeu 21
Db 110 ACTAAGTTTGGATTTTTCGATTTCTCTATGAGAAGCAGGACACACTTTTGAAGCTTCTC 169
Qy 22 IleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPhe 41
Db 170 ATTCTGTCATGGTGGTGTATATCTTCTCCACTGCTGTGTTGCTGCTGCTGAGATT 229
Qy 42 GluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAla 61
Db 230 GAAAGTGTATCCATGATGTTGATCCGTACTTTAATATCCGACTACAGGTTCTTGCT 289
Qy 62 GluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeuGly 81
Db 290 GAGGAGGGTTTTTATAAATCCATACTGGTTTGTATGACCGAGCTGTGTACCTTTGGGA 349

QY 82 ArgilleleGlyThrIleTyProGlyLeuMetIleThrSerAlaAlaIleTyHis 101
 Db 350 CGAATCAITGGAGAACAAATTAACCCAGGTTAATGATCACCTCTGCTGCAATCTACCAT 409
 QY 102 ValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaPro 121
 Db 410 GTACTCCATTTTTTCCACATACACATCCACATTCGGATGCTGTGTGTCTCTGGCCCT 469
 QY 122 LeuPheSerSerPheThrSer-IleValThrTyTrpLeu-LeuThrIleGluLeuLysAspA 141
 Db 470 CTCTCTCTCTCTCTCCACAC-CATCGTCAGGTACCA-CCTTACCAAGAGCTCAAGGATG 527
 QY 141 laGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValProGlyTyTrIleSerArgS 161
 Db 528 CAGGGCTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 587
 QY 161 exValAlaGlySerTyTrpAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyTr 181
 Db 588 CTGTGGCTGGCTCTCTATGATATGATGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 647
 QY 181 yMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeuA 201
 Db 648 ACATGTGGATCAAGGAGCAAGAGACTGGTTCATCTGTGGGAGCTAAGTGTGCCCTTG 707
 QY 201 latyrPheTyTrMetValSerSerTrpGlyGlyTyTrValPheLeuIleAsnLeuIleProL 221
 Db 708 CTATTTCTACATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 767
 QY 221 euHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyValAlaTyR 241
 Db 768 TCCACCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 827
 QY 241 yThrValTyCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPheGln 260
 Db 828 GTACTGTACTCTCTGGGCACTATCTCTCTAT-CCAGATCTCTTGTGGTTCAG 886
 QY 261 ProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnIleHi 280
 Db 887 CTGTCTCTCTATCAGACACATGGCAGC-CTTTGGGCTTGTGCTCTGCTGCTGCTGCTGCTGCT 945
 QY 280 sAlaPheValAspTyLeuArgSerTyLeuAsnProGlnPheGlnPheGluValLeuPheAr 300
 Db 946 TGCTTTGTGGATTACTGCGCAGCAAGTGAATCCACAACTTTGAAGTCTTTTCG 1005
 QY 300 gSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuTh 320
 Db 1006 GAGCGTCATCTCTGTAGGCTTTGCTCTCTCACCCTGGGAGCTCTCTCATGCTGAC 1065
 QY 320 rGlyLysIleSerProTrpThrGlyArgPheTyTrSerLeuLeuAspProSerTyAlaTy 340
 Db 1066 AGGAAAAATATCTCCCTGGACGGGGCTTTCTACTCGCTGGATCCCTCTTATGCTAA 1125
 QY 340 sAsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTy 360
 Db 1126 GAACACATCCCATCATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1185
 QY 360 rTyPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyTyCysPheSe 380
 Db 1186 CTATTTTGACCTGACGCT 1245
 QY 380 rAsnLeuSerAspAlaArgIlePheIleMetTyTrGlyValThrSerMetTyPheSe 400
 Db 1246 CAACCTGCTGATGCTCCCGGATTTTATCATCATGATGATGATGATGATGATGATGATGATGATGAT 1305
 QY 400 rAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerGlyTy 420
 Db 1306 AGCTGTAAATGGTGGCTCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1364
 QY 420 leGlyValSerGlnValLeuSerThrTyTrMetLysAsnLeuAspIleSerArgProAspL 440
 Db 1365 TTGAGTCTCCAGGCTCTCTCCACATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1424
 QY 440 ystLysSerLysGlnAspSerThrTyTrProIleLysIle-GluValAlaSerGly 459

Db 1425 AGAAGAGCAGAGAGCAACAGGATCCACCTACCTCTATTAAAG-TGAAGTGGCAAGTGGG 1483
 QY 460 MetIleLeuValMetAlaPhePheLeuIleThrTyTrPheHisSerThrTyValThr 479
 Db 1484 ATGATACTGGTTCATGGCTTTCTTTCTCATCACTACACCTTTCACTCACTGGGTGACC 1543
 QY 480 SerGluAlaTyTrSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArg 499
 Db 1544 AGTGAAGGCT 1603
 QY 500 IleIlePheAspAspPheArgGluAlaTyTrTrpLeuArgHisAsnThrProGluAsp 519
 Db 1604 ATCATATTTGATGACTCTCCGAGAAGCATATTTATGGCTTCGTATAATCTCCAGAGAT 1663
 QY 520 AlaLysValMetSerTrpTrpAspTyTrGlyTyTrGlnIleThrAlaMetAlaAsnArgThr 539
 Db 1664 GCGAAGGCTCATGCTCTCTGGGATTTATGGCTATCAGATTACAGCTATGGCAACCGAACA 1723
 QY 540 IleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMet 559
 Db 1724 ATTTAGTGGACAATAACACATGGAATAATACCCATATTTCTCGAGTAGGGCAGCAATG 1783
 QY 560 AlaSerThrGluGluLysAlaTyTrGluIleMetArgGluLeuAspValSerTyValLeu 579
 Db 1784 GCGTCCACAGAGAAAGCCCTATGAGATCATGAGGAGCTCGATGTGAGCTATGTGTGCTG 1843
 QY 580 ValIlePheGlyGlyLeuThrGlyTyTrSerSerAspAspIleAsnLysPheLeuTrpMet 599
 Db 1844 GTCATTTTGGAGGCTCTCACTGGTATTCTCTGATGATATCAACAAGTTCTTTGGATG 1903
 QY 600 ValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyTrThr 619
 Db 1904 GTCCGATTTGGAGGAGCAGATACAGCAACATATACAGGAGATGACTATTATTACT 1963
 QY 620 ProThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeuMet 639
 Db 1964 CCACTCGGGAGTTCGCTGTGACCGTGAAGTTCCTCAAGTCTCTCACTGCTCATG 2023
 QY 640 TyTrLysMetCysTyTrTyArgPheGlyGlnValTyThrGluAlaLysArgProGly 659
 Db 2024 TACAAGATGTGTACTATGCTTTGGACAGGTTTACACAGAGCCCAAGCGTCTCCAGGC 2083
 QY 660 PheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 679
 Db 2084 TTTGACCGTCTCGAAATGCTGAGATTGGATTAAGACTTTTGAGTCTGATGCTCTGGAG 2143
 QY 680 Glu 680
 Db 2144 GAA 2146
 RESULT 11
 BC020965
 LOCUS
 DEFINITION Homo sapiens integral membrane protein 1, mRNA (cdna clone MGC:9042 IMAGE:3891543), complete cds.
 ACCESSION BC020965
 VERSION BC020965.2
 KEYWORDS GIC:34190969
 SOURCE MGC.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2516)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, J., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,

929	Db	CCATGCCCTTTGGGATTACCTCGCGACGAAGTTGAATCCACAAATTTGAAGTCTCTTT	988
299	Qy	eAqSerValIleSerLeuValIcIyPheValLeuLeuThrValGlyAlaLeuLeuMetIe	319
989	Db	CCGGAGCGTCATCTCTCTGGTAGGCTTTGTCTCTCACCGTGGGAGCTCTCTCATGCT	1048
319	Qy	uThrGlyIysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAl	339
1049	Db	GACAGAAAAATATCTCCCTCGACCGGGCGTTTCTACTCGCTGGATCCCTCTTATGC	1108
339	Qy	alySerAsnIleProIleAlaSerValSerGluHisGlnProThrThrTyrSerSe	359
1109	Db	TAGAAACACATCCCATCATTTCTGTCTGTGACATCAGCCACACACTGGTCTCTC	1168
359	Qy	rTyrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCyaph	379
1169	Db	ATACTATTTTGACCTGCAGCTCTCTCTTCTATGTTTCAGTTGGCTCTATTAAGTCT	1228
379	Qy	eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPh	399
1229	Db	TAGCAACCTGTCTGATGCCCGGATTTTATCATCATGTGATGTGTGACCGCATGTACTT	1288
399	Qy	eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG	419
1289	Db	TTTCAGCTGTAATGGTGGCTTAATGCTAGTGTGGCACCTGTTATGTG-CAITTCCTCTG	1347
419	Qy	lyIleGlyValSerGlnValLeuSerThrTyrMetIysAsnLeuAspIleSerArgProA	439
1348	Db	GCATTCGAGTCTCCAGGTCTGTCCACATACATGAAGAATCTGGACATAAGTCGTCCAG	1407
439	Qy	spIysIysSerIysGlnGlnAspSerThrTyrProIleIysIle-GluValAlaSer	458
1408	Db	ACAAGAAGACCAAGACACAGAGTTCACCTACCTATTAAAGAA-TGAGTGGCGACT	1466
459	Qy	GlyMetIleLeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTrpVal	478
1467	Db	GGGATGATACGTGGTCATGGCTTCTTCTTCATCACCTACACCTTTCATTCAACCTGGGTG	1526
479	Qy	ThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer	498
1527	Db	ACCAGTAGAGCCCTACTCTCTCCGTCATTGTGACTATCTGCCCGCTGGTGGGATGGCACT	1586
499	Qy	ArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu	518
1587	Db	AGGATCATATTTGATGACTTCCGAGAAGCATATTAATGGCTTCGTATATATCCACAGAG	1646
519	Qy	AspAlaIysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg	538
1647	Db	GATCGGAAGTCATGCTCCTGGTGGGATATGGCTATCAGATTACAGATTATGGCAACCCGA	1706
539	Qy	ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla	558
1707	Db	ACAATTTTAGTGACATACACATGGAAATATACCCATATTTCTCGATAGGCGCAGGCA	1766
559	Qy	MetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal	578
1767	Db	ATGGCGTCCACAGAGGAAAAGCCTATGAGATCATGAGGAGGTCCGATGTCAGCTATGTG	1826
579	Qy	LeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspIleAsnIysPheLeuTrp	598
1827	Db	CTGGTCAITTTTGGAGGCGCTCACITGGGTATTCCTCTGATGATATCAACAGTTTCTTGG	1886
599	Qy	MetValArgIleGlyGlySerThrAspThrGlyIysHisIleIysGluAsnAspTyrTyr	618
1887	Db	ATGTCGCGGATTTGGAGGCGGACACAGATACAGGCAAAACATATCAAGGAGAAATGACTATTAT	1946
619	Qy	ThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeu	638
1947	Db	ACTCCAACTGGGGAGTTCCGTGTGACCGGTGAAGTTCTCCAGTGTGCTCTCACTGCCCTC	2006
639	Qy	MetTyrIysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProPro	658
2007	Db	ATGTACAAAGATGTGTACTATCGTTTGGACAGGTTTACAGAAAGCAAGCGTCTCCCA	2066

Db 523 GGATCCAGGGGCTGGGCTTCTGCTGCTGCCCATGATTGCTGTAGTTCCTCGATATATCTC 582
 Qy 159 rArgSerValAlaGlySerTyrAspAsnGluGlyAlaIlePheCysMetLeuLeuTh 179
 Db 583 CGGATCTGCTGGCTGCTCTATGATATGATGAGGAGTTCCTATCTTTTGGCATGCTACTCAC 642
 Qy 179 rTyrTyrMetTrpIleValAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAl 199
 Db 643 CTACTACATGTGGATCAGGAGTAAGACCTGGTTCATCTGTTGGCAGCTAAGTGTGC 702
 Qy 199 aLeuAlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuLeuLeuIle 219
 Db 703 CCTTGGCTTATCTACATGGTCTCGTCACTGGGAGGTTATGTGTCTCGTCACTCACTAAT 762
 Qy 219 eProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAl 239
 Db 763 TCTCTCCAGTCTCTGGTGTGATGCTCCAGGCGGTTTCTCTCACCGGATCTATGGGC 822
 Qy 239 aTyrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyP 259
 Db 823 CTACTGTACTGTCTTACTGCTGGGCACCTATATCTTCTAT-GCAGATCTCCTTTGTGGTT 881
 Qy 259 heGlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGln 278
 Db 882 TCCAGCCTGCTCTTTCATCAGACACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAG 940
 Qy 279 IleHisAlaPheValAspTyrIleuArgSerIleuAsnProGlnGlnPheGluValLeu 298
 Db 941 ATCCATGCTTTGTGGATTAATCTGGCAGCAAGTTGAATCCACAACTTTGAAGTTCTT 1000
 Qy 299 PheArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMet 318
 Db 1001 TTCGGAGGGTCACTCTCTGGTAGGCTTTGTCTTCTCACCGTGGAGTCTCTCTCATG 1060
 Qy 319 LeuThrGlyLysIleSer-ProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyr 338
 Db 1061 CTGACAGGAAAAATATCTCCCTGGAGCGGGCTTTCTACTCGCTGCTGGATCCCTCTTAT 1120
 Qy 339 AlaLysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrTrpSer 358
 Db 1121 GCTAAGACACATCCCATCATCTCTCTGCTGCTGAGCATCAGCCCACTGGTCTC 1180
 Qy 359 SerTyrTyrPheAspLeuLeuLeuValPheMetPheProValGlyLeuTyrTyrCys 378
 Db 1181 TCATATAATTTGACCTGCAGCTCTCTCTCTATGTTTCAGTGGCTCTATTACTGC 1240
 Qy 379 PheSerAsnLeuSerAspAlaArgPheIleIleMetTyrGlyValThrSerMetTyr 398
 Db 1241 TTTAGCAACCTGCTGATGCTCCGGATTTTATCATCATGATGATGATGATGATGATGAT 1300
 Qy 399 PheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSe 418
 Db 1301 TTTTCAGCTGTAAATGGTGGCTTAATGCTAGTGTGGCAGCTGTTATGTG-CATTCTCTC 1359
 Qy 418 rGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgPr 438
 Db 1360 TGGCATTGAGTCTCCAGTCTGCTGCACATGATGAAGAACTGGACATAGTCTGCTC 1419
 Qy 438 oAspLysLysSerLysGlnAspSerThrTyrProIleLysIle-GluValAlaAs 458
 Db 1420 AGACAAG 1478
 Qy 458 erGlyMetIleLeuValMetAlaPheLeuIleThrTyrPheHisSerThrTrp 478
 Db 1479 GTGGGATGATCTGGTCACTGGCTTCTTCTCTCATCAGCTACCTTCATTCACCTGGG 1538
 Qy 478 alThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGly 498
 Db 1539 TGACCAAGTGGAGCTACTCTTCTCCGCTCACTTGTCTGCTGCTGCTGCTGCTGCTGCT 1598
 Qy 498 erArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrPro 518
 Db 1599 GTAGGATCATATTTGATGACTCTCCGAGAGAGCATATATTGGCTTCTGTCATTAATCTCAG 1658

Qy 518 luApAlaLysValMetSerTyrTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnA 538
 Db 1659 AGGATGCGAAGGTCATGCTCTGGTGGGATTCGCTATCAGATTACAGATTGCGCAACCC 1718
 Qy 538 rGThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerA:GValGlyGlnA 558
 Db 1719 GAACAATTTAGTGGACAATAACACATGGATTAATACCCATATTTCTCGAGTAGGCAGG 1778
 Qy 558 laMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrV 578
 Db 1779 CAATGGCGTCCACAGAGAGAAAGCCCTATCAGATCATGAGGAGCTCGATGTGAGTATG 1838
 Qy 578 alLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuT 598
 Db 1839 TGCTGGTCATTTTGGAGGCTCTCATCGGTATTCTCTGATGATATCAACAAGTTCTTT 1898
 Qy 598 rpMetValArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrT 618
 Db 1899 GGATGGTCCGATTTGGAGGAGGACACATACAGGCAACATATCAAGGAGAAATGACTATT 1958
 Qy 618 yrThrProThrGlyLysPheArgValAspArgGluGlySerProValLeuLeuAsnCysL 638
 Db 1959 ATACTCAACTGGGAGTTCGTTGGACCGTGAAGTTCCTCCAGTCTGCTCACTGCC 2018
 Qy 638 euMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProp 658
 Db 2019 TCATGTACAGATGTGTACTATCTGCTTTGGACAGGTTTACACAGAGCCCAAGCGTCC 2078
 Qy 658 rcGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspVal 678
 Db 2079 CAGGCTTTGACCGTTCGGAATGCTGAGATGGGAATAAAGACTTTGAGCTTGATGTC 2138
 Qy 678 euGluGlu 680
 Db 2139 TGGAGGAA 2146

RESULT 13

HSM809254 2898 bp mRNA linear PRI 30-AUG-2003
 Homo sapiens mRNA; cDNA DKEZp779N2434 (from clone DKEZp779N2434).

DEFINITION BX649102
 VERSION BX649102.1 GI:34368274

KEYWORDS Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2898) Mews,H.W., Weil,B., Amid,C., Osanger,A.,

Lauber,J., Bahr,A., Fobo,G., Han,M. and Wiemann,S.
 The German Human cDNA Consortium

CONSRM Direct Submission
 JOURNAL Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764

COMMENT Neuberberg, GERMANY
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by Qigen (Hilden/Germany) within the cDNA sequencing

consortium of the German Genome Project.
 This clone (DKEZp779N2434) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further

information about the clone and the sequencing project is available
 at http://mips.gsf.de/proj/cDNA/.

Location/Qualifiers
 1..2898

/organism="Homo sapiens"
 /mol_type="mRNA"

/db_xref="taxon:9606"
 /clone="DKEZp779N2434"

/tissue_type="liver"
 /clone_lib="779 (synonym: hnccl). Vector pSport1_sfi; host

DH10B; sites sfiA + sfiB"

FEATURES
 source

/dev_stages="fetal"

polyA_signal
polyA_site

ORIGIN

Alignment Scores:

Pred. No.: 4,15e-220 Length: 2898
 Score: 220.00 Matches: 671
 Percent Similarity: 97.39% Conservative: 0
 Best Local Similarity: 97.39% Mismatches: 9
 Query Match: 31.21% Indels: 18
 DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x HSM809254 (1-2898)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 186 ATGACTAAGTTGGATTTTTCGGATTGCTCTATGAGAGCGGACACATTTTGAAGCTT 245
 QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 DB 246 CTCATTCTGTCAATGGCTGCTGTATTATCTCTCCACTGCTGTTGCTGCTGCTGAGA 305
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 DB 306 TTGAAAGTGTATCCAGTGGTTGATCCGTTTAAATTAATCCGACTACCAAGTTCCTG 365
 QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTyrTyrProLeu 80
 DB 366 GCTGAGGAGGGGTTTATAAATTCATAAAGTGGTTTGATGACCGAGCTGGTACCCCTTTG 425
 QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 DB 426 GGACGATCATGAGGAGACAAATTAATCCAGGTTTAAATGATCACTCTCTGCTGCATCTAC 485
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal-CysValPheLeuAl 120
 DB 486 CATGTACTTCCATTTTTCACATCACCATCGACATTCGGAATGC-CTGTGTGTTCTCTGCG 544
 QY 120 aProLeuPheSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLys 139
 DB 545 CCCT 602
 QY 140 AspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSer 159
 DB 603 GATCAGGGGCTGGGCTCTCTGCTGCTCCATGCTGTAGTTCCTGATATATCTCC 662
 QY 160 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 179
 DB 663 CGATCTGGCTGGCTCTCTATGATATGAAGGATTCCTATCTTTGCAATGCTACTCACC 722
 QY 180 TyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAla 199
 DB 723 TACTACATGTGGATCAAGGAGTAAAGACATGGTTCCATCTGTTGGGAGCTAAGTGTGCG 782
 QY 200 LeuAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIle 219
 DB 783 CTTCCTTATTTACATGGTCTGCTGATGGGAGGTTATGTTTCTCTGATCACTTAAT 842
 QY 220 ProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAla 239
 DB 843 CCTCTCCAGCTCTCTGCTGCTGATGCTCACAGGCGCTTCTCTCACCGGATCTATGTGCG 902
 QY 240 TyrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
 DB 903 TACTGTACTGTCTTACTGCTGGGCACTATCTTCTAT-CCAGATCTCTCTTGTGGGTTT 961
 QY 259 eGlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279
 DB 962 CCAGCCCTGTCTCTTTCATCAGAGCACATGGCAGC-CTTTGGGGTCTTTTGTCTCTCTCCACA 1020
 QY 279 leHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuP 299

DB 1021 TCCATGCTTTTGTGGATTACCTGGCAGCAAGTTGAATCCACAACAATTTGAAGTTCTTT 1080
 QY 299 heArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetL 319
 DB 1081 TCCGAGAGCGTCATCTCTGTGTAGGCTTTGTCTCTTCCACGCTGGAGCTCTCTCTCATGC 1140
 QY 319 eunThrGlyValIleSerProTyrThrGlyArgPhe-TyrSerLeuLeuAspProSerTyr 338
 DB 1141 TGACAGAAAATAATCTCCCTGGACGGGGCTCT-CTACTGCTGCTGGATCCCTCTTAT 1199
 QY 339 AlaLysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrSer 358
 DB 1200 GCTAAGAACAACATCCCATCTCTCTGTGTGTGAGCATCAGCCACACACCTGGTCC 1259
 QY 359 SerTyrTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrCys 378
 DB 1260 TCATATCTATTTTGACCTGGAGCTCTCTCTCTTTCATGTTTCCAGTTGGCTCTATATGTC 1319
 QY 379 PheSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyr 398
 DB 1320 TTTAGCAACCTCTCTGATGCCCGGATTTTATATCATCATGTATGTTGACCAAGCATGTAC 1379
 QY 399 PheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSe 418
 DB 1380 TTTTTCAGCTGTAAATGGTGGCTCTAATGCTAGTTGGACCTGTATGTG-CATCTCTC 1438
 QY 418 rGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgPr 438
 DB 1439 TGGCAATTGGAGCTCTCCAGGCTCTCTCCACATACATGAAGAATCTGGACATAAGTCGTCC 1498
 QY 438 oAspLysLysSerLysGlnAspSerThrTyrProIleLysIle-GluValAlaLys 458
 DB 1499 AGCACAGAGACAGAGACACAGGATTCACCTACCTATTAAGAA-TGAAGTGGCAA 1557
 QY 458 er-GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTyr 477
 DB 1558 GTGGGATA-ATACTGGTCATGGCTTTCTTCTCTCACTACACTACACTTTCAATCAACCTGG 1616
 QY 478 ValThrSerGluAlaTyr-SerSerProSerIleValLeuSerAlaArgGlyGlyAspGly 497
 DB 1617 GTGACACAGTGGGCTCTCTCTCTCCCTCAATGCTATCTGCGCGTGGTGGGATGCG 1676
 QY 498 SerArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrPro 517
 DB 1677 AGTAGATCATATTTGATGACTTCCGAGAGCATATATTGGCTTCGTCTGCTAATACTCCA 1736
 QY 518 GluAspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsn 537
 DB 1737 GAGGATGCCAGAGTCATGCTCTGGTGGATTAATGGCTATCAGATTACAGCTATGGCAAAC 1796
 QY 538 ArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGln 557
 DB 1797 CGAACCAATTTTGTAGTGACATAACACATGGAATAATACCCATATTTCTCGAGTAGGCGAG 1856
 QY 558 AlaMetAlaSerThrGluLysLysAlaTyrGluIleMetArgGluLeuAspValSerTyr 577
 DB 1857 GCATGGCTCCACAGAGGAAAAGCCTATGAGATCATGAGGAGCTTGTATGTCAGCTAT 1916
 QY 578 ValLeuValIlePheGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeu 597
 DB 1917 GTGCTGTCTATTTTGGAGGCTCTCTCTGATGATTCCTCTGATGATCAACAAGTTTCTT 1976
 QY 598 TrpMetValArgIleGlyLysSerThrAspThrGlyLysHisIleLysGluAsnAspTyr 617
 DB 1977 TGGATGGTCCGATTTGGAGGAGGACACATACAGGCAACATATCAAGGAGATGACTAT 2036
 QY 618 TyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCys 637
 DB 2037 TATACTCCAACTGGGAGTTCGCTGGACCGGTGAAGGTTCTCCAGTGTGCTCAAGTCC 2096
 QY 638 LeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgPro 657
 DB 2097 CTCATGTACAAAGATGTGTACTATCTCTCTTGGACAGGTTTACACAGAAAGCCAGCTCT 2156

QY 658 ProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspVal 677
 Db 2157 CCAGGCTTTGACCGGTGCGAAATGCTGAGATTGGGAATAAAGACTTTTGAGCTTTGATGTC 2216
 QY 678 LeuGluGlu 680
 Db 2217 CTGGAGGAA 2225

RESULT 14
 AK130548
 LOCUS Homo sapiens cDNA FLJ27038 fis, clone SLV08213, highly similar to
 DEFINITION Oligosaccharyl transferase STT3 subunit homolog.
 ACCESSION AK130548
 VERSION AK130548.1 GI:34527376
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 Kanehori, K., Ishibashi, T., Chiba, Y., Fujimori, K., Hiraoka, S.,
 Tanai, H., Watanabe, S., Ishida, S., Ono, Y., Hotuta, T., Watanabe, M.,
 Suzuki, Y., Hata, H., Nakagawa, K., Mizuno, S., Morinaga, M.,
 Kawamura, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,
 Nishikawa, T., Sugiyama, A., Kawakami, B., Nagai, K., Isogai, T. and
 Sugano, S.

TITLE NEDO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1796)

AUTHORS Sugano, S. and Suzuki, Y.

TITLE Direct Submission

JOURNAL

Submitted (31-JUL-2003) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Laboratory of Genome Structure, Human Genome
 Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
 (E-mail: fldnag@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
 Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; CDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction and 5'-end one pass sequencing: Institute of Medical
 Science, University of Tokyo, Laboratory of Genome Structure, Human
 Genome Center; 3'-end one pass sequencing: RAB; clone selection for
 full insert sequencing: RAB and Helix Research Institute.

FEATURES

source

1..1796
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="SLV08213"
 /tissue types="salivary gland"
 /clone_lib="SLV"
 /note="Cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:

Pred. No.: 6,95e-191 Length: 1796
 Score: 192.00 Matches: 305
 Percent Similarity: 98.07% Conservative: 0
 Best Local Similarity: 98.07% Mismatches: 3
 Query Match: 27.23% Indels: 6
 Gaps: 5

US-10-028-384-12 (1-705) x AK130548 (1-1796)

QY 373 ValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleLeuMetTyr 392

Db 233 GTTGGCTTATTTACTGCTTTAGCAACCTGCTCGATCCCGGATTTTATCATCATGAT 292

QY 393 GlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeuAlaPro 412

Db 293 GGTGTGACCAGCATGTACTTTTTCAGCTGTAAATGGTGGCTTAATGCTAGTGTGGCACT 352

QY 413 ValMetSer-IleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAs 432
 Db 353 GTTATGTG-CATTCTCTCTGGCATTTGGAGTCTCCAGGTGCTGTCCACATACATGAAGA 411

QY 432 nLeuAspIleSerArgProAspLysSerLysValSerGlnGlnAspSerThrTyrProI 452
 Db 412 TCTGGACATAAGTCGTCACAGCAAGAGAGCAAGAGCAAGAGGATTCACCTACCCAT 471

QY 452 elysIle-GluValAlaSerGlyMetIleLeuValMetAlaPhePheLeuIleThrTyr 472
 Db 472 TAAGAA-TGAAGTGGCAAGTGGCATGATCTGGTCATGCTTCTTCTCATCACCTACA 530

QY 472 hPheHisSerThrTrpValThrSerGluAlaTyrSerSerProSerIleValLeuSerA 492
 Db 531 CTTTTCATTCAACCTGGGTGACGACGAGGCGCTACTCTTCTCGTCCATTGACTATCTG 590

QY 492 laArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyrTyrPL 512
 Db 591 CCCGTGGTGGGATGGCAGTAGGATCATATTGATGACTTCCGAGACGATATTATTGGC 650

QY 512 euArgHisAsnThrProGluAspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnI 532
 Db 651 TTTCGTCAATAATACTCCAGAGGATGCGAAGGTCAATGCTCTGTGGGATTATGGCTATCAGA 710

QY 532 leThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisI 552
 Db 711 TTACGACTATGGCAACCGACAACTTTTAGTGGACAATAACACATGGGAATTAATACCCATA 770

QY 552 leSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgG 572
 Db 771 TTTCGTAGTAGGGGAGGAGGATGGCGTCCACAGAGGAAAAAGCCCTATCAGATCATGAGG 830

QY 572 luLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspA 592
 Db 831 AGCTCGATGTGAGTATGCTGTGCTCATTTTGGAGGCTCACCTGGGTATTCCTCTGTATG 890

QY 592 spIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHisI 612
 Db 891 ATATCAACAAGTTCTTTGGATGGTCCGATTTGGAGGAGCACAGATACAGGCAACATA 950

QY 612 leLysGluAsnAspTyrTyrThrProThrGlyGluPheArgValAspArgGluLysSerP 632
 Db 951 TCAAGGAGATGACTATTATTACTCAACTGGGGAGTTCCCGTGGACCGTGAAGGTCTC 1010

QY 632 roValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPhe-GlyGlnValTyr 651
 Db 1011 CAGTCTGCTCAACTGCTCATGTACAAAGATGTGTACTATCGCTG-TGGACAGGTTTAC 1059

QY 652 ThrGluAlaLysArgProProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLys 671
 Db 1070 ACAGAAGCCAAAGCGTCTCCAGGCTTTGACCGTGTCCGAAATGTGTGAGATTGGGAATAA 1129

QY 672 AspPheGluLeuAspValLeuGluGlu 680
 Db 1130 GACTTTGAGCTTTGATGCTCGGAGGAA 1156

RESULT 15
 HUNTMTC
 LOCUS Homo sapiens transmembrane protein (TMC) mRNA linear PRI 13-FEB-2002
 DEFINITION L47337
 ACCESSION L47337
 VERSION L47337.1 GI:18654193
 KEYWORDS transmembrane protein.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2293)
 AUTHORS Bellacosa, A. B. and Testa, J. J. R.
 TITLE Isolation, characterization and mapping to human chromosome
 11q24-25 of a gene, TMC, encoding a highly conserved putative
 transmembrane protein

JOURNAL Unpublished Location/Qualifiers
 FEATURES source
 1..2293
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /map="11924-25"
 /cell_type="mesothelioma"
 /clone_lib="ddc11"
 1..2293
 /gene="TMC"
 117..2234
 /gene="TMC"
 /note="putative"
 /codon_start=1
 /product="transmembrane protein"
 /protein_id="AA077539.1"
 /db_xref="GI:18654134"
 /translation="MTKFGFLRLSYEKQDTLLKLILLSMAVLSFSTRFLAVLRPESV
 IHEDPYFNRTFRUSEEGFYKHNWFDRAWYPLIGTIYFGIMTSAAIYHV
 LHPFHTIDIRNVCVFLAPLFSFTTIVYHLTKELKADAGLLAAAMIAVVPYISR
 SVAGSYNEGIAIFCMLLTYIMIKAVKTGSIQWAAKACALAYFVWVSGWGVFLINL
 IPLHVLMLTGRFSHRIYVAYCTVYCLGTILSMQISFVGPQVPLSSHMAAFGVFL
 COIHAYDYLRSKLNPQRPVLRSVISLVGVLLTVGALLMTGKISPTWGRFVSL
 DPSYAKNIPILIASVESHPTWSSVYFQLQLLVFMFVGLYCSNLSDAIRFLIMY
 GTSYKFSAMVRLMLVLARVCLIGISQVLSVLMNLSIAPDKSKKQDSTY
 PIKEVASGMILVNAFLITYTTFHSTWTSSEYSSPISLARDGDSRIIFDDPREA
 YYURHNTPSDAKVMWYGYQITAMANRTILVDNNTWNTTHISRGQANASTBEKA
 YEIMRELDVSLVIFEGLTGYSDDINKFLMVRIGGSTDTGKHIRENDYVTPTEF
 RVDRGSPVLLNCLMYKCYRFGQVYTEAKRPFGDPRVNAEIGNKDFELDLVEAY
 TTEHVLVRIYKVKLDNRGLSRT"

ORIGIN
 Alignment Scores:
 Pred. No.: 1,72e-184 Length: 2293
 Score: 186.00 Matches: 672
 Percent Similarity: 97.67% Conservative: 0
 Best Local Similarity: 97.67% Mismatches: 8
 Query Match: 26.38% Indels: 16
 DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x HUMTMC (1-2293)

QY 1 MetThrIysPheGlyPheLeuArgLeuSerTyGluLysGlnAspThrLeuLeuLysLeu 20
 Db 117 ATGACTAAGATTGGATTTCGCGATTGCTCTATGAGAACGAGCACACATTTTGAAGCTT 176
 QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 Db 177 CTCATTCTGTCATAGCGTGGCTGTATATCTCTCCACCTCGTCTGCTGCTCTGAGA 236
 QY 41 PheGluSerValIleIleGluPheAspProTyPheAsnTyArgThrThrArgPheLeu 60
 Db 237 TTTGAAAGTGTATCCATGATGTTGATCCGCTACTTAAATATCGGACTACCAAGTTCCTG 296
 QY 61 Ala-GluGluGlyPheTyIysPheHisAsnTTPheAspAspArgAlaTrpTyProle 80
 Db 297 TC-TGAGAGAGGGGTTTTATAATCCATACTGGTTTGATGACCGGCTGTACCCCTTT 355
 QY 80 uGlyArgIleIleGlyThrIleTyProGlyLeuMetIleThrSerAlaAlaIleTy 100
 Db 356 GGGACGATCATTTGAGAGAAACAATTTACCCAGGTTTTAAATGATCACCCTCTGCTGCAATCTA 415
 QY 100 rHisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAl 120
 Db 416 CCATGTACTCAATTTTCCACATCCATCCATCGACATTCGGAATGCTGTGTGCTCTGCGC 475
 QY 120 aProLeuPheSerSerPheThrSer-IleValThrTyIleLeuLeuThrIysGluLeuLys 139
 Db 476 CCCT 533
 QY 140 AspAlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValProGlyTyIleSer 159

Db 534 GATGCAGGGGCTGGGCTTCTTCTGCTGCCATGATTGCTGTAGTTCCTGGATATATCTCC 593
 QY 160 ArgSerValAlaGlySerTyArgPheGlnGluGlyIleAlaIlePheCysMetLeuLeuThr 179
 Db 594 CGATCTGTGGCTGGCTCTTATGATTAATGAAGGGATTGCCATCTTTTGCATGCTACTCACC 653
 QY 180 TyTyMetTrpIleLysAlaValLysThrGlySerIleCysTTPAlaAlaLysCysAla 199
 Db 654 TACTACATGTGGATCAAGGCACGTAAAGACTGGTTCATCTGTTGGGACGCTAAGTGTGCC 713
 QY 200 LeuAlaTyPheTyMetValSerSerTrpGlyGlyTyValPheLeuLeuLeuLeuLeu 219
 Db 714 CTTTGGCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTGTTCTTCATCACTTAATT 773
 QY 220 ProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyValAla 239
 Db 774 CTTCTCCAGCTCTCGTCTGTATGCTCACAGGCCGTTTCTCTCACCGGATCTATGTGCC 833
 QY 240 TyCysThrValTyCysLeuGlyThrIleLeuSerArg-GlnLeuSerPheValGlyPhe 259
 Db 834 TACTGTACTGTTTACTGCTGGGCACTATCTTCTAT-GCAGATCTCTCTTGTGGGTTT 892
 QY 259 eGlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279
 Db 893 CCAAGCTCTCTTCTATCAGACACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAGA 951
 QY 279 IeHisAlaPheValAspTyLeuArgSerLysLeuAsnProGlnGlnPheGluValLeu 299
 Db 952 TCCATGCCCTTTGTGGATTACCTGGCAGCAAGATTGAATCCCAACAATTTGAAGTCTTT 1011
 QY 299 heArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyValAlaLeuLeuMet 319
 Db 1012 TCCGAGCGTCACTCTCTGTGAGCTTTTGTCTCTCACCGTGGAGCTCTCTCTATGC 1071
 QY 319 euThrGlyLysLeuSerProThrThrGlyArgPheTySerLeuLeuAspProSerTyArg 339
 Db 1072 TGACAGAGAAATATCTCCTGGACGGGCGGTTCTACTCGCTCTGGATCCCTCTTATG 1131
 QY 339 lAlysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerS 359
 Db 1132 CTANGAACACATCCCATCATGCTTCTGTGTCTGAGCATCAGCCACACACCTGGTCT 1191
 QY 359 erTyTyThrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyTyCysP 379
 Db 1192 CATACTATTTTACCTGCGAGCTCCCTCGCTTCATGTTTCCAGTTGGGCTCTATTACTGCT 1251
 QY 379 heSerAsnLeuSerAspAlaArgIlePheIleIleMetTyGlyValThrSerMetTyP 399
 Db 1252 TTAGCAACCTGTCTGATGCCGGATTTTATCATCATGATATGGTGTGACCGCATGTACT 1311
 QY 399 heSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSer 418
 Db 1312 TTTCACTGTAAATGGTGGCTCTAAATGCTAGTGTGGACCTGTATGTG-CATTCTCTCT 1370
 QY 419 GlyIleGlyValSerGlnValLeuSerThrTyMetLysAsnLeuAspIleSerArgPro 438
 Db 1371 GGATTTGGAGTCTCCAGGTCGTGTCACATACATGAAGATCTGGACATAGTCGCCCA 1430
 QY 439 AspLysSerLysLysGlnAspSerThrTyProIleLysIle-GluValAlaIle 458
 Db 1431 GACAAGAGAGCAAGAACAGAGATTCACCTACCTACCTATTAAGAA-TGAAGTGGCAAG 1489
 QY 458 rGlyMetIleLeuValMetAlaPhePheLeuIleTyThrPheHisSerThrTrpVa 478
 Db 1490 TGGATGTACTGTGTCTGCTTCTCTCATCACCTACACCTTTTCTTCACTCACTGGGT 1549
 QY 478 lThrSerGluAlaTySerSerProSerIleValLeuSerAlaArgGly-GlyAspGlyS 498
 Db 1550 GACCAAGTGGGCTCTCTTCTCCGTCATGTTACTTCTGCCCGTGA-TGGGGATGGCA 1608
 QY 498 exArgIleIlePheAspPheArgGluAlaTyTyThrIleArgHisAsnThrProG 518
 Db 1609 GTAGGATCATATTTGATGACTTCCGAGAGCATATTTATGCTTCCTCGTCAATAATCTCCAG 1668

```
Qy 518 luAspAlaLysValMetSerTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnA 538
Db 1669 AGGATGCGAAGGTCAATGCTGGTGGATTATGGCTATCAGATTACAGCTATGGCAACC 1728
Qy 538 rgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnA 558
Db 1729 GAACAAATTTTGTAGTGACAAATAACACATGGATAATATCCCATATTTCTCGAGTAGGGCAGG 1788
Qy 558 laMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrV 578
Db 1789 CAATGGCGCTCCACAGAGGAAAGCCCTATGAGATCATGAGGAGCTCGATGTGAGCTATG 1848
Qy 578 alleuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeut 598
Db 1849 TGCTGGTCATTTTGGAGGCCCTCACTGGGTATTCCTCTGATGATATCAACAAGTTTCCTT 1908
Qy 598 rpMetValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrT 618
Db 1909 GGATGGTCCGGATTGGAGGGAGCACACATACAGGCAACATATCAAGGAGAAATGACTATT 1968
Qy 618 yrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysL 638
Db 1969 ATACTCCAACCTGGGAGTTCCGTGTGGACCGTGAAGGTTCCTCCAGTCTCTCAACTGCC 2028
Qy 638 euMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProp 658
Db 2029 TCATGTACAGATGTGTACTATCGCTTTGGACAGGTTTACAGAGGCCAAGCGTCCTC 2088
Qy 658 roGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspVal 678
Db 2089 CAGGCTTTGACCGTGTCCGAAATGCTGAGATTGGGAATAAGACTTTGAGCTTGATGTCC 2148
Qy 678 euGluGlu 680
Db 2149 TGGAGGAA 2156
```

Search completed: December 14, 2004, 22:48:51
Job time : 7872.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 : Search time 812.278 Seconds
(without alignments)
4556.130 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 705

Sequence: 1 MTKFGFLRLSEKQDTLLKL:.....HMLVRIYKVDLDRGLSRT 705

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+pin.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131645_6790/app_query.fasta_1.3740
-DB=N_Geneseq_23Sep04 -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=us10028384 @CGN 1.1.1986 @runat_14122004_131645_6790 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database: N_Geneseq_23Sep04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	705	100.0	2472	10	ADD94793 Human ITM
2	705	100.0	2472	10	ADH28844 Human chr
3	226	32.1	2760	5	ABV24502 Human pro
4	152	21.6	3094	10	ADD94791 Mouse ITM
5	146	20.7	440	6	ABK27622 Human col
6	140	19.9	455	6	ABL37867 Human col

7	128	18.2	487	3	AAA16208	Aaa16208 Human col
8	127	18.0	1371	12	ADP28508	Adp28508 Human sec
9	126	17.9	575	3	AAC98825	Aac98825 Human pan
10	102	14.5	414	3	AAC00273	Aac00273 Human sec
11	102	14.5	487	6	ABK39617	Abk39617 cDNA enco
12	102	14.5	487	8	ACA11946	Aca11946 Human lun
13	102	14.5	487	8	ACA03132	Aca03132 Lung carc
14	102	14.5	487	10	ADH47174	Adh47174 Human lun
15	101	14.3	307	5	AAF68440	Aaf68440 Human lun
16	101	14.3	307	6	ABK39233	Abk39233 DNA enco
17	101	14.3	307	6	ABK38351	Abk38351 cDNA enco
18	101	14.3	307	8	ACA11562	Aca11562 Human lun
19	101	14.3	307	8	ACA10680	Aca10680 Human lun
20	101	14.3	307	8	ABX99631	Abx99631 Lung carc
21	101	14.3	307	8	ACA02748	Aca02748 Lung carc
22	101	14.3	307	10	ADH46790	Adh46790 Human lun
23	101	14.3	307	10	ADH45894	Adh45894 Human lun
24	101	14.3	307	12	ADE71646	Ade71646 Human lun
25	94	13.3	419	8	ABX45128	Abx45128 Bovine ES
26	81	11.5	596	6	ABN73253	Abn73253 Bovine em
27	81	11.5	596	6	ABN73343	Abn73343 Bovine em
28	80	11.3	476	9	ACH34952	Ach34952 Human end
29	65	9.2	307	5	AAF68653	Aaf68653 Human lun
30	65	9.2	307	6	ABK38564	Abk38564 cDNA enco
31	65	9.2	307	8	ACA10893	Aca10893 Human lun
32	65	9.2	307	8	ABX98444	Abx98444 Lung carc
33	65	9.2	307	10	ADH46107	Adh46107 Human lun
34	65	9.2	307	12	ADE71859	Ade71859 Human lun
35	59	8.4	573	6	ABQ55493	Abq55493 Human ova
36	54	7.7	526	5	AAH83346	Aah83346 Human ova
37	53	7.5	516	4	AA524717	Aas24717 Human ova
38	42	6.0	261	8	ABX35963	Abx35963 Bovine ES
39	42	6.0	382	8	ABX41175	Abx41175 Bovine ES
40	40	5.7	473	6	ABQ58556	Abq58556 Human col
41	40	5.7	515	6	ABL38310	Ab138310 Human col
42	40	5.7	665	6	ABQ59387	Abq59387 Human col
43	39	5.5	2855	4	ABL02795	Ab102795 Drosophil
44	39	5.5	6153	4	ABL02794	Ab102794 Drosophil
45	38	5.4	243	6	ABV86822	Abv86822 Human col

ALIGNMENTS

RESULT 1
ADD94793
ID ADD94793 standard; DNA; 2472 BP.

XX

AC ADD94793;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human ITM1 gene sequence.

XX

XX source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cytostatic; immunosuppressive; antineoplastic; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
KW gene; ds; ITM1.

XX Homo sapiens.

XX

PN WO2003054008-A2.

XX

PD 03-JUL-2003.

XX

PF 18-DEC-2002; 2002WO-CA001967.

XX

PR 20-DEC-2001; 2001US-00028384.

XX

PA (COMP-) COMPATIGENE INC.

Db 1607 ATATTGATGACTCCGAGAGCATATTATTGGCTTCGTCAATACTCCAGAGGATGCG 1666
 QY 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 Db 1667 AAGTCATGTCCTGGTGGGATATGCTATCAGATACAGCTATGCGAAACCGAACAAATT 1726
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1727 TTAGTGGACAATAAACATGGAATAATACCCATATTTCTCGAGTAGGGCAGGCAATGCG 1786
 QY 561 SerThrGluGluValAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1787 TCCACAGAGGAAAGCCATATGATCATGAGGAGCTCGATGTCAGTATGCTGCTGTC 1846
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 Db 1847 ATTTTGGAGGCTCACTGGGTATTCCTCTGATGATATCAACAAGTTTCTTTGGATGCTC 1906
 QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1907 CGATTTGGAGGAGCAGATACAGGCAACAAATATCAAGGAGATGACTATTACTCCA 1966
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1967 ACTGGGAGTTCGCTGGACCGTGAAGGTTCTCCAGTGTCTCAACTGGCTCATGTAC 2026
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysAspProGlyPhe 660
 Db 2027 AAGATGTGTACTATCGCTTTGGACAGGTTTACAGAGCCAGCGTCTCCAGGCTTT 2086
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 2087 GACCGTGTCCGAAATGCTGAGATTGGCAATAAAGACTTTGAGCTTGATGTCTCGAGAA 2146
 QY 681 GlyTyrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2147 GCGTATACACAGAACTTGGCTGTGTCAGGATATACAGGATATACAGGACCTGATATCGA 2206
 QY 701 GlyLysSerArgThr 705
 Db 2207 GGCTTGTCAAGGACA 2221

RESULT 2

ID ADH28844 standard; DNA; 2472 BP.

XX ADH28844;

XX 11-MAR-2004 (first entry)

DE Human chronic myelogenous leukaemia (CML) gene marker #112.

XX ds; chronic phase chronic myelogenous leukaemia; CP-CML;
 KW Blast crisis CML; BC-CML; human; chronic myelogenous leukaemia;
 XX Gene marker.

XX Homo sapiens.

XX US2003104426-A1.

XX 05-JUN-2003.

XX 14-JUN-2002; 2002US-00171581.

XX 18-JUN-2001; 2001US-0298914P.

XX (LINS// LINSLEY P S.

XX (MAOM// MAO M.

XX (DAIH// DAI H.

XX (REYI// HE Y.

XX (RADI// RADICH J P.

XX Linsley PS, Mao M, Dai H, He Y, Radich JP;

XX

XX WPI; 2003-787046/74.

XX Classifying cell sample as chronic phase chronic myelogenous leukemia or
 PT blast crisis chronic myelogenous leukemia by detecting difference in
 PT expression of genes corresponding to the markers such as X15415, U89436.

XX Disclosure; SEQ ID NO 112; 31pp; English.

XX The invention relates to a method of classifying a cell sample as chronic
 CC phase chronic myelogenous leukaemia (CP-CML) or blast crisis CML (BC-
 CC CML). The method is useful for classifying a sample as CP-CML or BC-CML.
 CC The present sequence represents a human chronic myelogenous leukemia
 CC (CML) gene marker used to distinguish blast crisis CML from chronic phase
 CC CML.

XX Sequence 2472 BP; 566 A; 568 C; 583 G; 755 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2472
 Score: 705.00 Matches: 705
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-12 (1-705) x ADH28844 (1-2472)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 Db 107 ATGACTAAGTTTGGATTTTTCGATTGTCCTATGAGAGCAGGACACACTTTTGAAGCTT 166
 QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 Db 167 CTCATCTGTCNAUGGCTGCTGATATCTCTCCACTCGTCTGTTGCTGCTCTGAGA 226
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 Db 227 TTTGAAGTGTATTCATCATGAGTTTTCATCCGCTACTTTTAAATATCGAGCTACAGGTTCTCTG 286
 QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTyrTyrProLeu 80
 Db 287 GCTGAGGAGGGGTTTATTAATTCATAACTGGTTTTCATGACCGAGCTGCTGACCTTTG 346
 QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 347 GGACGAATCAITGGAGGAACAATTATCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 406
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db 407 CATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGTCGTGTGTGTTCTGGCC 466
 QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 Db 467 CCTCTCTTCTCCTCCTCCATCCATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 526
 QY 141 AlaGlyValGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
 Db 527 GCAGGGGCTGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 586
 QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 Db 587 TCTGTGGTGGCTCCTATGATAATGAAGGATTCCTCATCTTTCATGCTGCTGCTGCTGCTG 646
 QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
 Db 647 TACATGTGGATCAAGGAGTAAGACTGGTTCATCTGTTGGGAGCTAGTGTGCTGCTGCT 706
 QY 201 AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuLeuPro 220
 Db 707 GCTTATTTCTACATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 766
 QY 221 LeuHisValLeuValLeuMetLeuThrThrGlyArgPheSerHisArgIleTyrValAlaTyr 240

767	DB	CTCCACGCTCTCGTGGCTGATGCTCACAGGCCGGTTTCTCTCACCGGATCTATGTGGCCCTAC	836
241	QY	CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln	260
827	DB	TGTAAGTGTACTGCTGGGTACTATACCTTTCTAGGCAGATCTCTTTGGTGTTCAG	886
251	QY	ProValIleuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis	280
887	DB	CTGTGCTTTTCATCAGACACATGGCAGGGTTGGGCTTTGGTCTCTGCCAGATCCAT	946
281	QY	AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnInPheGluValLeuPheArg	300
947	DB	GCCTTTGTGGATTACCTGCGCACCAAGTTGAAATCCACAACAATTTGAAGTTCCTTTCCGG	1006
301	QY	SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr	320
1007	DB	AGCGTCATCTCTCTGGTAGGCTTTGCTTCTCACCGTGGGAGCTCTCTCATGCTGACA	1066
321	QY	GlyIlysIleSerProThrPheGlyArgPheTyrSerLeuLeuAppProSerTyrAlaLys	340
1067	DB	GGAAAAATATCTCCCTGGAGCGGGCGTTTCTACTCAGTGTGGATGCCCTTTATGCTAAG	1126
341	QY	AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr	360
1127	DB	AACAACATCCCATCATTTGCTGTGCTGAGCATCAGGCCACACAACCTGGTCTCTCATAC	1186
361	QY	TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer	380
1187	DB	TATTTGACCTGAGCTCTCTGCTCTTCATGTTTCAGTTGGCTCTATTACTGCTTAGC	1246
381	QY	AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer	400
1247	DB	AACCTGTCTGATGCCCGGATTTTATCATCATGTATGTGTGACACACATGTAATTTCA	1306
401	QY	AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle	420
1307	DB	GCTGTAAATGGTGGCTCTAAATGCTAGTGTGGCACCTGTTATGAGCATTCCTCTGGCAT	1366
421	QY	GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys	440
1367	DB	GGAGTCTCCACGGTGTCTCCATACATACATGAGANCTGGACATAGTGCACACACAG	1426
441	QY	LysSerIlysGlnInAspSerThrTyrProIleLysIleGluValAlaSerGlyMet	460
1427	DB	AAGAGCAAGAACCAACAGGATCCACCTACCTATTAAGATTGAAGTGGCAAGTGGGATG	1486
461	QY	IleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThrSer	480
1487	DB	ATACTGTGTCATGCGCTTCTTCCTCATCACCCTACACCTTTCACTCGGTGACCAAT	1546
481	QY	GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle	500
1547	DB	GAGCCCTACTCTCTCCGTCCATGTACTATCTGCCGTGGTGGGATGGGAGTAGATC	1606
501	QY	IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla	520
1607	DB	ATATTTGATGACTCCGGAAGCATATTATTGGCTTCGTCAATAATCTCCAGAGGATGG	1666
521	QY	LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle	540
1667	DB	AAGGTGATGTCCTGGTGGGATTATGGCTATCAGATTACAGCTATGGCAACCCGACAA	1726
541	QY	LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla	560
1727	DB	TTAGTGACAAATAACACATGGAATAATCCCATATTTCTCAGTAGGGCAGGCAATGGCG	1786
561	QY	SerThrGluIlysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal	580
1787	DB	TCCACAGAGGAAAAGCCTATGAGATCATGAGGGAGCTCGATGTACAGTATGCTGTGTC	1846
581	QY	IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal	600

[illegible]

CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX

SQ Sequence 2760 BP; 688 A; 645 C; 658 G; 769 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4, 1e-224 Length: 2760
 Score: 226.00 Matches: 674
 Percent Similarity: 97.9% Conservative: 0
 Best Local Similarity: 97.9% Mismatches: 6
 Query Match: 32.06% Indels: 14
 DB: 5 Gaps: 0

US-10-028-384-12 (1-705) x ABV24502 (1-2760)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 105 ATGACTAAGTTGGGATTTTGGGATTTGCTTATGCTTCCATGAGAGCAGGACACATTTTGAAGCTT 164
 QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu-PheAlaValLeuAr 40
 DB 165 CTCATTCGTCAATGGCTGCTGATATATCCCTTCCTCCACTGCTGTTTGTGCTGCTGAG 224
 QY 40 gPheGluSerValLeuHisGluPheAspProTyrPheAsnTyrArgThrArgPheLe 60
 DB 225 ATTTGAAAGTGTATCCATGATGATTTGATCCGTACTTTAATATATCGGACTACCAAGTTCTT 284
 QY 60 uAlaGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrProLe 80
 DB 285 GGCTGAGAGGGGTTTTATAATTTCCATGCTGTTGATGACCGAGCTGGTACCCCTTT 344
 QY 80 uGlyArgLeuLeuGlyThrLeuTyrProGlyLeuMetLeuThrSerAlaAlaLeuTy 100
 DB 345 GGGACGAATATCGAGGACAAATTTACCCAGGTTTAAATGATACCTTCCTGCTGCATCTA 404
 QY 100 rHisValLeuHis-PhePheHisLeuThrLeuAspLeuArgAsnValCysValPheLeuA 120
 DB 405 CCATGACTCCATTTTTTCCACATCACCATCGACATCGGAATGCTGTGTGTCTCTGG 464
 QY 120 laProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLy 139
 DB 465 CCCCTCTCTCTCTCTCTACAC-CATGTCACGTTACCA-CCTTACCAAGAGTCAA 522
 QY 139 sAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSe 159
 DB 523 GGATGCAAGGGGCTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 582
 QY 159 rArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuTh 179
 DB 583 CCGATCTGTGGTGGCTCTCTATGATATGAGGGATGGCATTTTGTGATGCTATCTAC 642
 QY 179 rTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAl 199
 DB 643 CTACTACATGTGGATCAAGCAGTAAAGACTGGTTCATCTGTTGGCAGCTAAGTGTGC 702
 QY 199 aLeuAlaTyrPheTyrMetValSerSerTrpGlyTyrValPheIleuLeuLeuL 219
 DB 703 CCTTGTCTATTTCTACATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 762
 QY 219 eProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAl 239
 DB 763 TCCTCTCCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 822
 QY 239 aTyrCysThrValTyrCysLeuGlyThrLeuLeuSerArg-GlnIleSerPheValGlyP 259
 DB 823 CTACTGTACTGTTTACTGCTGGGCACTATACCTTTCTAT-GCAGATCTCTCTTTTGGGTT 881
 QY 259 heGlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGln 278
 DB 882 TCCAGGCTGTCTTTTCATCAGACACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAG 940

QY 279 IleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeu 298
 DB 941 ATCCATGCCCTTTGGATTTACCTGCGCAGCAAGTTGAATCCACACAAATTTGAAGTTCTT 1000
 QY 299 PheArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMet 318
 DB 1001 TTCGGAGCGTCACTCTCTGGTAGGCTTTGTCCTTCTCACCGTGGAGCTCTCTCTCATG 1060
 QY 319 LeuThrGlyLysIleSerProTrpThrGlyArgPheTyr-SerLeuLeuAspProSerTyr 338
 DB 1061 CTGACAGGAAAAATATCTCCCTGACGGGGCTTTCTACTCGTGTGATCCCTCTTAT 1120
 QY 339 AlaLysAsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSer 358
 DB 1121 GCTAAGAACAAACATCCCATCTGTTCTGTGTCTGAGCATCAGCCCAACCTGGTCC 1180
 QY 359 SerTyrTyr-PheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCys 378
 DB 1181 TCATACTATTTGACCTGCGCTCTCTGCTCTTCATGTTTCCAGTTGGCCTCTATTACTGC 1240
 QY 379 PheSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyr 398
 DB 1241 TTTAGCAACCTGCTGATGCCGGATTTTATCATCATGTATGTTGTGACCAAGCATGTAC 1300
 QY 399 PheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSe 418
 DB 1301 TTTTCAGCTGTAATGGTGGCTCTAATGCTAGTGTGGACCTGTTATGTG-CATTCTCTTC 1359
 QY 418 rGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgPr 438
 DB 1360 TGCGCATTTGAGTCTCCAGGTCTCTCCACATACATGAAGATCTGGACATAAGTCGTC 1419
 QY 438 oAspLysLysSerLysGlnLysAspSerThrTyrProIleLysIle-GluValAlaAs 458
 DB 1420 AGACAGAAGAGCAAGAACAAACAGATTCCACCTACCTATTAAGAA-TGAAGTGGCAA 1478
 QY 458 exGlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpV 478
 DB 1479 GTGGGATGATATCTGGTCTGCTTCTTCTCATCACCTACACCTTTTCATTCAACCTGG 1538
 QY 478 alThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlys 498
 DB 1539 TGACCAAGTGGAGCTTACTCTCTCCGTCATGTTCTGCTGCCGCTGGTGGGATGGCA 1598
 QY 498 exArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProG 518
 DB 1599 GTAGGATCATATTTGATGACTTCCGAGAGCATATATTGGCTTCGTCATTAATCTCCAG 1658
 QY 518 luAspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnA 538
 DB 1659 AGGATGCGAAGGTCATGTCCTGGTGGGATTTATGGCTATCAGATTACAGCTATGGCAACC 1718
 QY 538 rGThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnA 558
 DB 1719 GBAACAATTTTATGGACAATAACATGGAATAATACCATATTTCTCGAGATGGGAGG 1778
 QY 558 laMetAlaSerThrGluGlyLysAlaTyrGluIleMetArgGluLeuAspValSerTyrV 578
 DB 1779 CAATGGCTCCACAGAGAGGAAAAAGCTATGAGATCATGAGGAGCTCGATGTCAGCTATG 1838
 QY 578 alLeuValIlePheGlyLysLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuT 598
 DB 1839 TGCTGGTCATTTTGGAGGCTCTACTGGTATTTCTCTGATGATATCAACAAGTTTCTTT 1898
 QY 598 rpMetValArgIleGlyLysSerThrAspThrGlyLysHisIleLysGluAsnAspTyrT 618
 DB 1899 GGATGCTCGGATTTGGAGGAGCAGACATACAGGCAACATATCAAGAGAAATGACTATT 1958
 QY 618 yrThrProThrGlyLysPheArgValAspArgGluGlySerProValLeuLeuAsnCysL 638
 DB 1959 ATATCCAACTGGGAGTTCGTTGGACCGTGAAGGTTCTCCAGTCTCTCAACTGCC 2018

QY 638 euMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProp 658
 Db 2019 TCATGTACAAGATGTTACTATCGCTTTGGACAGGTTTACACAGAGGCAAGCGTCTC 2078
 QY 658 roGlyPheAspArgValArgAsnAlaGluLeuGlyAsnLysAspPheGluLeuAspVal 678
 Db 2079 CAGGCTTTGACCGGTGTCGGAATGCTGAGATTGGGAATAAGACTTTTGAGCTTGATGCC 2138
 QY 678 euGluGlu 680
 Db 2139 TGGAGGA 2146

RESULT 4
 ADD94791
 ID ADD94791 standard; DNA; 3094 BP.
 XX
 AC ADD94791;
 DT 29-JAN-2004 (first entry)
 XX
 DE Mouse ITM1 gene sequence.
 XX
 KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cystostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 KW murine; Gene; ds; ITM1.
 XX
 OS Mus musculus.
 XX
 PN WO2003054008-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 18-DEC-2002; 2002WO-CA001967.
 XX
 PR 20-DEC-2001; 2001US-00028384.
 XX
 PA (COMP-) COMPATIGENE INC.
 XX
 PI Perreault C, McBride K;
 XX
 DR WPI; 2003-559122/52.
 DR P-PSDB; ADD94792.
 XX

New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.

Disclosure; SEQ ID NO 9; 66pp; English.

This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC) - associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestine cancer, sarcomas, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer, or haematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the mouse ITM1 gene which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from GenBank.

XX
 SQ Sequence 3094 BP; 758 A; 651 C; 703 G; 982 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,78e-147 Length: 3094
 Score: 152.00 Matches: 667
 Percent Similarity: 97.37% Conservative: 0
 Best Local Similarity: 97.37% Mismatches: 9
 Query Match: 21.56% Indels: 18
 DB: 10 Gaps: 0
 US-10-028-384-12 (1-705) x ADD94791 (1-3094)
 QY 5 GlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeuLeuLysLeuSer 24
 Db 124 GGATTTTTCGATTGTCCTATGAGAGGAGGACACACTTCTAAAGCTTCTCATCTGTCG 183
 QY 25 MetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerVal 44
 Db 184 ATGGCTGCTGTGTTATCTTTTCTACTCTTTTGTGCTGAGATTGAAAGTGC 243
 QY 45 IleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGly 64
 Db 244 ATCCATGAGTTTATCGTACTTAAATATCGACTACCCGGTTTCTGGCTGAGGAGGG 303
 QY 65 PheTyrLysPheHisAsnTrpPheAspArgAlaTyrTyrProLeuGlyArgLeu 84
 Db 304 TTTTNTAATTCATTAACCTGGTTTGATGACCGGGCTTGTACCTTTGGGCGGATCAT 363
 QY 85 GlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHis 104
 Db 364 GGAGGAACAATTTACCCAGGTTTAAATGATCACTTCTGCTGCAATCTACCATGTATCCAT 423
 QY 105 PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSer 124
 Db 424 TCTTTCATATCATTATGACATTCGGAATGCTGTGTTTCTTGGGCCCTCTTCTTCTCC 483
 QY 125 SerPheThrSer-IleValThrTyrIleu-LeuThrLysGluLeuLysAspAlaGlyAla 144
 Db 484 TCTTTCACCAC-CATCGTTACGTACCA-CCTTACCAAGAGCTCAAGGATCAGAGGCTG 541
 QY 144 LysLeuAlaAlaAlaMetIleAlaValValProGlyTyrIleSerArgSerValAlaG 164
 Db 542 GGCTTCTTCTGCTGCCATGATGCTGTAGTTTCTGGGTATATTTCTGCACTGTAGCTG 601
 QY 164 LysTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetTrpI 184
 Db 602 GCTCTATGATAATGAAGGAATGCTATCTTTTGCATGCTGCTTACTTACTACATGTGA 661
 QY 184 LysAlaValLysThrGlySerIleCys-TpAlaAlaLysCysAlaLeuAlaTyr-Phe 203
 Db 662 TCAAGGAGTGAAGACTGGTTTCCATCTA-TTGGGCTGCAAGTGTGGCTTCTGCTTATTC 720
 QY 204 TyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuLeuProLeuHisVal 223
 Db 721 TACATGGTCTCTTCATGGGAGGCTATGTTTCTCTGATCAACTGATTCCTCTACATGTC 780
 QY 224 LeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrVal 243
 Db 781 CTGGTCTAATGCTGACAGGCGTTTCTTCTACCGGATCTAGTAGCTTACTTACTGTT 840
 QY 244 TyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPheGlnProValLe 263
 Db 841 TACTGCTGGGACCATTCTTCTAT-GCAGATTTCCTTTGTGGTTTCCAGCCGCTCT 899
 QY 263 uSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnIleHisAlaPheV 283
 Db 900 TCCATCAGAACACATGCGAGC-CTTTGGAGTGTGTTGTTCTGTGAGTCCATGCTTTCG 958
 QY 283 AlaSerTyrLeuArgSerLysLeuAsnProGlnInPheGluValLeuPheArgSerValI 303
 Db 959 TAGATTACCTGGGCGAGCAAGTTGAATCCACAGCAATTCGAAGTTCTTTTCCGGAGTGT 1018

QY 303 leSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThrGlyLysI 323
Db 1019 TCTCCCTGGTTGGCTTTGCTCCTCCTCACTGTGGGAGCTCTCCCATGCTAACAGGAA 1078
QY 323 leSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLysAsnI 343
Db 1079 TTTCTCCCTGGACAGGGGTTTCTACTCTCTGTGGATCCCTCTTATGCTAAGAA 1138
QY 343 leProIleAlaSerValSerGluHisGlnProThrThrTrpSerTyrTyrPheA 363
Db 1139 TTCCCATATTGCACTCTTTCTGAGCACAGCCACACCTGCTCTTCTACTATT 1198
QY 363 spLeuGlnLeuLeuValPheTetPheProValGlyLeuTyrTyrCysPheSerAsnLeuS 383
Db 1199 ATCTACAGCTCTGCTCTTCTGATTTTCCAGTTGGCTCTTATCTACTGCTTTAGCA 1258
QY 383 exAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSerAlaValM 403
Db 1259 CTGATGCTCGGATTTTATCATCATGATGATGCTGACCATGATCTACTTTTACGCTG 1318
QY 403 etValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerGlyIleGlyVal 422
Db 1319 TGGTGGCTTAATGCTGATGATGGACCTGTTATGTG-CATCTCTTCTGGCATGGTGT 1377
QY 423 SerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLysLysSer 442
Db 1378 TCCAGGTGCTGCTCACATATATGAANAATCTGCACATAGTCGCCACAGCAAGAGAGC 1437
QY 443 LysLysGlnGlnAspSerThrTyrProIleLysIle-GluValAlaSerGlyMetIleLe 462
Db 1438 AAGAAGCAACAGATCTACTTACCTTACCTTAAAGAA-TGAGGTGGCGAGTGGGATGAT 1496
QY 462 uValMetAlaPhePheLeuIleThrTyrPheHisSerThrTyrValThrSerGluAl 482
Db 1497 GGTATGCTTTTCTCATCACCCTACACCTTTCATTCGACTTGGGTGACCATGAGC 1556
QY 482 aTyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArgIleLePhe 502
Db 1557 CTATTCTTCTCCCTCCATGCTACTGCTGCTGGTGGGAGTGGATGATCAATTTT 1616
QY 502 eAspAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProGluAspAlaLysVa 522
Db 1617 TGATGATCTCCGAGAGCGTATTATGCTCCGTCACATATCTCCAGGATGCAAAAGT 1676
QY 522 MetSerTrpTrpAspTyrGlyGlnIleThrAlaMetAlaAsnArgThrIleLeuVa 542
Db 1677 CATGTCATGCTGGGATTTGCTACCAATTTACTGCAATGGCAATCGCAATTTTGT 1736
QY 542 lAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerTh 562
Db 1737 GGACATTAACACATGGATATATACCATTTCTCGAGTAGGGCAGGCAATGGCATCCAC 1796
QY 562 rGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValIleValIlePhe 582
Db 1797 AGAAGAAAGGCTATGAATCATGAGGAGCTTGATGTCAGCTATGCTGTGTTGTCA 1856
QY 582 eGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArgI 602
Db 1857 TGGAGGCTTACTGGGTATCTTCGGATGATATACCAAGTTCTTTGGATGGTCCGGAT 1916
QY 602 eGlyGlySerThrAsp-ThrGlyLys-HisIleLysGluAsnAspTyrTyrThrProThr 621
Db 1917 TGGAGGAAGCAGAG-GACAGG-AAGACACATTAAGGAGATGATCTACTATCTCTACT 1974
QY 622 GlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLys 641
Db 1975 GGGGATTCGGTTGATGCTGAGGGTCTCCGGTGTGCTCACTGCCATTATGACAAA 2034
QY 642 MetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPheAsp 661
Db 2035 ATGTGTACTACCGCTTTTGGGAGGTCTACAGAGCAAGCGTCCACAGGTTTGAC 2094
QY 662 ArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680

Db 2095 CGTGTTCGAATGCTGAGATTGGTATATAAGACTTTGAGCTTGATCTCTGGAGAA 2151

RESULT 5

ABK27622/c
ID ABK27622 standard; cDNA; 440 BP.
XX
AC ABK27622;
XX
DT 09-APR-2002 (first entry)
XX
DE Human colon cancer expressed sequence tag, Seq ID no 59.
XX
KW Human; colon cancer; T cell expansion; tumour; EST; gene; ss;
KW expressed sequence tag.

XX Homo sapiens.

XX WO200196390-A2.

XX 20-DEC-2001.

XX 08-JUN-2001; 2001WO-US018577.

XX 09-JUN-2000; 2000US-0210821P.

XX 18-DEC-2000; 2000US-0256571P.

XX 10-MAY-2001; 2001US-0290240P.

XX (CORI-) CORIYA CORP.

XX Jiang Y, Hepler WT, Clapper JD, Wang A, Secrist H;

XX WPI; 2002-139708/18.

XX Novel isolated polynucleotide encoding a polypeptide comprising a portion of colon tumor protein, useful for detection, diagnosis and therapy of human colon cancer.

XX Claim 1; Page 167; 220pp; English.

XX The invention relates to an isolated polynucleotide (I) encoding a polypeptide (II) comprising at least a portion of a colon tumour protein. (I), (II) and antibody (III) to (II) are useful for determining the presence of a cancer in a patient. (I), (II) or antigen presenting cells expressing (I) is useful for stimulating and/or expanding T cells specific for a tumour protein, by contacting T cells with (I), (II) or antigen-presenting cells that express (I), under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells. (I), (II), or antigen presenting cells that express (I) are useful for treating colon cancer in a patient by incubating CD4+ and/or CD8+ T cells isolated from a patient with (I), (II) or antigen presenting cells that express (II) such that T cells proliferate, and administering to the patient an effective amount of the proliferated T cells, thus inhibiting the development of a cancer in the patient. (I) or (II) is useful in vaccines and pharmaceutical compositions for prevention and treatment of colon malignancies and for the diagnosis and monitoring of such cancers. (I), (II) or (III) is useful for detection, diagnosis and/or therapy of human colon cancer. (I) is useful as a probe or primer for nucleic acid hybridisation, and in the design and preparation of ribozyme molecules for inhibiting expression of (II) in tumour cells. ABK27564-ABK27807 represent novel human colon cancer coding sequences and primers of the invention

XX Sequence 440 BP; 117 A; 119 C; 84 G; 120 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,02e-141 Length: 440
Score: 146.00 Matches: 145
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.71% Indels: 0
DB: Gaps: 0

US-10-028-384-12 (1-705) x ABK27622 (1-440)

QY 494 GlyGlyAspGlySerArgIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArg 513
 Db 439 GGTGGGATGCGATAGATCATATTGATGATTTCCGAGACATATTATTGGCTTCT 380
 QY 514 HisAsnThrProGluAspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThr 533
 Db 379 CATATACTCCAGAGGATGCGAGGTCATGCTCTGGTGGGATTATGGCTATCAGATTACA 320
 QY 534 AlaMetAlaAsnArgThrIleLeuValAspAsnThrTrpAsnThrHisIleSer 553
 Db 319 GCTATGGCAACCCGACATTTTATGTGACATATACATGGAATATATCCCATTTCT 260
 QY 554 ArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeu 573
 Db 259 CCAAGTAGGCGCAGCAATGGCGTCCACAGAGGAAAAGCCATGAGATCATGAGGAGTCT 200
 QY 574 AspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIle 593
 Db 199 GATGTCAGTATGCTGGTTCATTTTGGAGGCTCCTGGGATTTCTCTGATGATTC 140
 QY 594 AsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHisIleLys 613
 Db 139 AACAAAGTTTCTTGGATGGTCCGATGGAGGAGGACACATACAGGCAACATATCAAG 80
 QY 614 GluAsnAspTyrTyrTrpThrGlyGluPheArgValAspArgGluGlySerProVal 633
 Db 79 GAGATGACTATATATCTCACTGGGAGTTCCTGGTGGAGCCGTGAGGTTCTCCAGTG 20
 QY 634 LeuLeuAsnCysLeuMet 639
 Db 19 CTGCTCAACTGCCCTCATG 2

RESULT 6

ABL37867
 ID ABL37867 standard; cDNA; 455 BP.
 XX AC ABL37867;
 XX DT 08-APR-2002 (first entry)
 XX DE Human colon tumour antigen polynucleotide SEQ ID NO:1456.
 XX DE Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
 XX DE Human tumour metastatic antigen; diagnosis; gene; ss.
 XX OS Homo sapiens.
 XX PN WO200196388-A2.
 XX PD 20-DEC-2001.
 XX PF 08-JUN-2001; 2001WO-US018557.
 XX PR 09-JUN-2000; 2000US-0210899P.
 XX PR 20-FEB-2001; 2001US-0270216P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Jiang Y, Harlocker SL, Secrist H;
 XX DR WPI; 2002-114514/15.
 XX PT Novel isolated colon tumor polynucleotide differentially expressed in
 XX PT colon tumor or colon metastatic tumor and polypeptides encoded by them,
 XX PT useful for inhibiting development of cancer in patient.
 XX PS Claim 1; SEQ ID NO 1456; 105pp; English.
 XX CC ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)
 XX CC which were isolated from human colon tumour and colon metastatic tumour
 XX CC cDNA libraries. (I) have cytostatic activity and can be used in vaccine

CC production. (I) can be used for stimulating and/or expanding T cells
 CC specific for a tumour protein on contact with the T cells. They are also
 CC useful for inhibiting the development of cancer in a patient. (I) can be
 CC used as probes or primers for nucleic acid hybridisation, for preparing
 CC mutant species primers, or primers for use in genetic constructions. (I)
 CC can be used in the diagnosis of a colon tumour
 XX SQ Sequence 455 BP; 123 A; 89 C; 120 G; 120 T; 0 U; 3 Other;

Alignment Scores:

Pred. No.: 1 82e-135 Length: 455
 Score: 140.00 Matches: 140
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 19.86% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-12 (1-705) x ABL37867 (1-455)

QY 490 LeuSerAlaArgGlyGlyAspGlySerArgIlePheAspAspPheArgGluAlaTyr 509
 Db 2 CTATCTGCCCGTGGTGGGATGGCAGTAGGATCATATTGATGACTTCCGAGAGCATAT 61
 QY 510 TyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrpAspTyrGly 529
 Db 62 TATTGGCTTCTGTCATAATATCTCCAGAGGATGCGAAGGTCAATGCTCTGGTGGATTATGGC 121
 QY 530 TyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsn 549
 Db 122 TATCAGATTACAGCTATGCGAACCCGAACTTTTAGTGGCAATAACATACATGGAATAAT 181
 QY 550 ThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIle 569
 Db 182 ACCATATTTCTCGAGTAGGCGAGCAATGGCGTCCACAGAGGAAAAAGCCTATGAGATC 241
 QY 570 MetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSer 589
 Db 242 ATGAGGGAGCTCGATGTCAGTATGCTGTGTCATTTTGGAGGCCCTCCTGGGTATTC 301
 QY 590 SerAspAspIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThrGly 609
 Db 302 TCTGATGATATCAACAAGTTCTTTGGATGGTCCGGATTGGAGGGAGCACAGATACAGGC 361
 QY 610 LysHisIleLysGluAsnAspTyrTrpThrProThrGlyGluPheArgValAspArgGlu 629
 Db 362 AAACATATCAGAGGAGATGACTATTATATCTCAACTGGGAGTTTCCGTGTGGACCGTGAA 421

RESULT 7

AAA16208
 ID AAA16208 standard; DNA; 487 BP.
 XX AC AAA16208;
 XX DT 14-JUN-2000 (first entry)
 XX DE Human colon cancer differentially expressed nucleotide sequence #213.
 XX DE Colon cancer; detect; differential expression; human; treatment;
 XX DE detect mutation; non-invasive diagnostic method; ds.
 XX OS Homo sapiens.
 XX PN WO200012702-A2.
 XX PD 09-MAR-2000.
 XX PF 30-AUG-1999; 99WO-US019424.
 XX PR 31-AUG-1998; 98US-0098639P.
 XX PR 27-JAN-1999; 99US-0117393P.
 XX PA (FARB) BAYER CORP.

PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
 PI Catano TJ, Dwivedi P, Ford DM, Lewis ME, Molino GA, Morahan JE,
 PI Schlegel R;
 XX WPI; 2000-256641/22.
 XX
 XX Novel nucleic acids and proteins for identifying therapeutic agents
 PT useful for treating and diagnosing cancer, especially colon cancer.
 XX
 XX Claim 16; Page 209; 345pp; English.
 XX
 XX This sequence represents a human nucleotide sequence which is
 CC differentially expressed in colon cancer cells compared to the expression
 CC levels in normal cells. The nucleotide sequence can be used as a source
 CC of primers and probes. The nucleotide sequence is useful for determining
 CC the phenotype of a cell by detecting the differential expression of the
 CC sequence relative to a normal cell. The probes derived from the sequence
 CC can also be used to determine the phenotype of cells in a sample. Probes
 CC and antibodies which hybridize to the nucleotide sequence can also be
 CC used to determine the phenotype of a cell. The primers are useful for
 CC detecting a mutation in a test nucleotide sequence and also for detecting
 CC cancer, preferably colon cancer. Antibodies against the protein encoded
 CC by the nucleotide sequence can also be used in a method to detect colon
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
 CC colon cancer at an early stage
 XX
 SQ Sequence 487 BP; 90 A; 140 C; 102 G; 155 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5,75e-123 Length: 487
 Score: 128.00 Matches: 128
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 18.16% Indels: 0
 DB: 3 Gaps: 0

US-10-028-384-12 (1-705) x AAA16208 (1-487)

QY 271 PheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeu 290
 Db 88 TTGGGGCTTTGGCTCTCCATGATCCATCCCTTTGGATACCTGGCAGCAGATTG 147
 QY 291 AsnProGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPheValLeu 310
 Db 148 AATCCACAACAAATTGAAGTTCTTTTCGGAGCGTCATCTCTCGTAGGCTTTGTCCT 207
 QY 311 LeuThrValGlyAlaLeuLeuMetLeuThrGlyLysIleSerProTyrThrGlyArgPhe 330
 Db 208 CTCACCGTGGGAGCTCTCTCATGCTGACAGAAATAATCTCCCTGGACGGCGGCTTC 267
 QY 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIlePheIleAlaSerValSer 350
 Db 268 TACTCACTGCTGGATCCCTCTATGCTTAAGAACACATCCCAATGCTTCTGTGCT 327
 QY 351 GluHisGlnProThrThrTrpSerSerTyrTyrPheAspLeuGlnLeuValPheMet 370
 Db 328 GAGCATCAGCCACCAACCTGGCTCTCATACATTTTACCTGCAGCTCCCTGCTTCATG 387
 QY 371 PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIle 390
 Db 388 TTTCAGTTGGCTCTATTACTGCTTTAGCAACCTGCTGATGCCCGGATTTTATCATC 447
 QY 391 MetTyrGlyValThrSerMetTyr 398
 Db 448 ATGTATGGTGTGACCAGCATGTAC 471

RESULT 8

ID ADP28508
 ID ADP28508 standard; DNA; 1371 BP.

XX

AC ADP28508;

XX

DT 12-AUG-2004 (first entry)

XX Human secreted protein encoding sequence SEQ ID #506.
 DE
 XX Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
 KW cancer; inflammatory; immune; ds; human secreted protein.
 XX
 OS Homo sapiens.
 XX WO2004035732-A2.
 PN
 XX 29-APR-2004.
 PD
 XX
 XX 28-AUG-2003; 2003WO-USQ26780.
 XX
 XX 29-AUG-2002; 2002US-0406576P.
 XX 29-AUG-2002; 2002US-0406579P.
 XX 29-AUG-2002; 2002US-0406585P.
 XX 29-AUG-2002; 2002US-0406588P.
 XX 29-AUG-2002; 2002US-0406608P.
 XX 29-AUG-2002; 2002US-0406611P.
 XX 29-AUG-2002; 2002US-0406612P.
 XX 29-AUG-2002; 2002US-0406616P.
 XX 29-AUG-2002; 2002US-0406640P.
 XX 29-AUG-2002; 2002US-0406642P.
 XX 29-AUG-2002; 2002US-0406646P.
 XX 29-AUG-2002; 2002US-0406653P.
 XX 29-AUG-2002; 2002US-0406655P.
 XX 29-AUG-2002; 2002US-0406666P.
 XX 17-SEP-2002; 2002US-0410946P.
 XX 17-SEP-2002; 2002US-0410947P.
 XX 17-SEP-2002; 2002US-0410948P.
 XX 17-SEP-2002; 2002US-0410949P.
 XX 17-SEP-2002; 2002US-0410953P.
 XX 17-SEP-2002; 2002US-0410957P.
 XX 17-SEP-2002; 2002US-0410958P.
 XX 17-SEP-2002; 2002US-0410959P.
 XX 17-SEP-2002; 2002US-0410960P.
 XX 17-SEP-2002; 2002US-0410961P.
 XX 17-SEP-2002; 2002US-0410962P.
 XX 17-SEP-2002; 2002US-0411019P.
 XX 17-SEP-2002; 2002US-0411022P.
 XX 17-SEP-2002; 2002US-0411023P.
 XX 17-SEP-2002; 2002US-0411024P.
 XX 17-SEP-2002; 2002US-0411032P.
 XX 17-SEP-2002; 2002US-0411035P.
 XX 17-SEP-2002; 2002US-0411037P.
 XX 17-SEP-2002; 2002US-0411041P.
 XX 17-SEP-2002; 2002US-0411045P.
 XX 17-SEP-2002; 2002US-0411046P.
 XX 17-SEP-2002; 2002US-0411048P.
 XX 17-SEP-2002; 2002US-0411052P.
 XX 17-SEP-2002; 2002US-0411055P.
 XX 17-SEP-2002; 2002US-0411073P.
 XX 17-SEP-2002; 2002US-0411082P.
 XX 17-SEP-2002; 2002US-0411101P.
 XX 17-SEP-2002; 2002US-0411111P.
 XX 18-APR-2003; 2003US-0463700P.
 XX 18-APR-2003; 2003US-0463708P.
 XX 18-APR-2003; 2003US-0463716P.
 XX 02-MAY-2003; 2003US-0463732P.
 XX 02-MAY-2003; 2003US-0467199P.
 XX 02-MAY-2003; 2003US-0467201P.
 XX 02-MAY-2003; 2003US-0467203P.
 XX 02-MAY-2003; 2003US-0467230P.
 XX 19-MAY-2003; 2003US-0471336P.
 XX 19-MAY-2003; 2003US-0471336P.
 XX 22-MAY-2003; 2003US-0472420P.
 XX 22-MAY-2003; 2003US-0472430P.
 XX 09-JUN-2003; 2003US-0476609P.
 XX 09-JUN-2003; 2003US-0476641P.
 XX 08-JUL-2003; 2003US-0485218P.
 XX 08-JUL-2003; 2003US-0485232P.
 XX 08-JUL-2003; 2003US-0485244P.

PR 08-JUL-2003; 2003US-0485325P.
 PR 14-JUL-2003; 2003US-0486446P.
 PR 14-JUL-2003; 2003US-0486480P.
 PR 15-JUL-2003; 2003US-0486891P.
 PR 15-JUL-2003; 2003US-0486960P.
 PR 08-AUG-2003; 2003US-0493341P.
 PR 08-AUG-2003; 2003US-0493370P.
 PR 08-AUG-2003; 2003US-0493573P.
 PR 08-AUG-2003; 2003US-0493577P.
 PA (FIVE-) FIVE PRIME THERAPEUTICS INC.
 XX
 PI Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
 PI Halenbeck RF, Huang MM, Kothakota S, Haishan L, Linnemann T;
 PI Pierce K, Wang Y, Wong JGP, Wu G, Zhang H;
 XX
 DR WPI, 2004-348438/32.
 XX
 XX New nucleic acid molecule for diagnosing, preventing or treating diseases
 PT such as proliferative (e.g. cancer), inflammatory, immune, metabolic,
 PT genetic, bacterial and viral diseases.
 XX
 PS Claim 1; SEQ ID NO 506; 428bp; English.
 XX
 CC The present invention relates to an isolated nucleic acid molecule
 CC encoding a polypeptide which is believed to be cytostatic,
 CC antiinflammatory, immunosuppressive, antibacterial and virucidal. The
 CC composition and methods are useful for diagnosing, preventing and
 CC treating diseases such as proliferative (e.g. cancer), inflammatory,
 CC immune, metabolic, genetic, bacterial and viral diseases. The present
 CC sequence represents a human secreted protein encoding sequence. The
 CC present sequence is available on WIPWEB and is not in the specification.
 XX
 SQ Sequence 1371 BP; 283 A; 350 C; 298 G; 440 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,69e-121 Length: 1371
 Score: 127.00 Matches: 408
 Percent Similarity: 97.61% Conservative: 0
 Best Local Similarity: 97.61% Mismatches: 6
 Query Match: 18.01% Indels: 10
 DB: 12 Gaps: 0

US-10-028-384-12 (1-705) x ADP28508 (1-1371)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 1 ATGACTAAGTTTGGATTTTGGCATTTGCTATGAGACAGCAGGACACATTTTGAAGCTT 60
 QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 DB 61 CTCATTCTGCAATGGCTGCTGTATTATCTCTCCACCTCGTCTGTGCTGCTGAGA 120
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 DB 121 TTTCAAAGTGTATCCATGATGTTTGTATCTCTTAAATATCGGACTACACAGGTTCTG 180
 QY 51 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
 DB 181 GCTGAGGAGGGGTTTATAAATCCATACTGTTTGTATGACCGAGCCTGTGACCTTTG 240
 QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 DB 241 GGACGAATCATTTGGAGGACAAATTTACCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 300
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 DB 301 CATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGTCTGTGTGTCTCTGCC 360
 QY 121 ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysA 140
 DB 361 CCTCTCTTCT 418

QY 140 sPlacIyAlacIyLeuLeuAlaAlaAlaMetIleAlaValProGlyTyrIleSerA 160
 DB 419 ATGCAGGGGCTGGGCTTCTTCTGCTGCCATGATTGCTGTAGTTCCTGGATATATCTCCC 478
 QY 160 rGSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrT 180
 DB 479 GATCTGTGGCTGGCTCTCTATGATATGAAGGATTGCCATCTTTTGATGCTACTCACT 538
 QY 180 yTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAla 200
 DB 539 ACTACATGGATCAAGCAGCAAGATGCTGTCATCTGTGGCAGCTAAAGTGTGCC 598
 QY 200 euAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIleP 220
 DB 599 TTGCTTATTCTACATGGTCTGTCATGGGAGGTTATGTCTCTGATCAACTTAATTC 658
 QY 220 rLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAla 240
 DB 659 CTCCTCACGCTCTGCTGCTGATGCTCAGACGCGGTTCTCTCACCGGATCTATGTGCC 718
 QY 240 yrcysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
 DB 719 ACTGTACTGTTTACTGCTGGGCACTATATTTCTAT-GCAGATCTCTTTTGGGTTTC 777
 QY 260 GlnProValLeuSerSerGlnHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279
 DB 778 CAGCCTGTCTCTTCTATCAGACACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAG 836
 QY 279 eHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuP 299
 DB 837 CCATGCTTTTGGATTACCTTGGCAGCAAGTTGAATCCACCAACAATTTGAAGTCTTTT 896
 QY 299 eArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuMetLe 319
 DB 897 CCGGAGCGCATCTCTCTGGTAGGCTTGTCTCTCACCGTGGAGCTCTCTCATGC- 955
 QY 319 uThrGlyLysIleSerProTyrThrGlyArgPheTyr-SerLeuLeuAspProSerTyrAl 339
 DB 956 -TG-GGAAAAATATCTCTCTGACGGGCGTTTCTACTCGCTGCTGGATCCCTCTTATGC 1013
 QY 339 alyAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSe 359
 DB 1014 TAAGAACACATCCCATCATGCTTCTGTCTGTAGCATCAGCCCAACACCTGGTCTCT 1073
 QY 359 rTyrTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysP 379
 DB 1074 ATACTATTGTACCTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1133
 QY 379 eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPh 399
 DB 1134 TAGCAACCTGTCTGATGCGCGGATTTTATCATCATGATGTGTGTGACCGCATGTACT 1193
 QY 399 eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMet 414
 DB 1194 TTCAGTGTAAATGGTGGCTGCTAAATGCTAGTGTGGCAGCTGTTATG 1239
 RESULT 9
 AAC98825
 ID AAC98825 standard; cDNA; 575 BP.
 XX
 AC AAC98825;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:53.
 XX
 KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection;
 KW diagnosis; identification; cytostatic; neuroprotective; nootropic;
 KW immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW immune system; muscular; reproductive; gastrointestinal; pulmonary;
 KW cardiovascular; renal; proliferative; ss.

OS Homo sapiens.
 XX WO200005320-A1.
 PN 21-SEP-2000.
 PD 08-MAR-2000; 2000WO-US005989.
 PF 12-MAR-1999; 99US-0124270P.
 PR (HUMA-) HUMAN GENOME SCI INC.
 XX PA Rosen CA, Ruben SM;
 XX PI WPI; 2000-579444/54.
 XX DR P-PSDB; AAB54060.
 DR
 XX
 PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition.
 XX
 PS Claim 1; Page 533; 1373pp; English.
 XX
 CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, nocotropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiac and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing in a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention
 XX
 SQ Sequence 575 BP; 117 A; 144 C; 137 G; 174 T; 0 U; 3 Other;

Alignment Scores:
 Pred. No.: 8,08e-121 Length: 575
 Score: 126.00 Matches: 126
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.87% Indels: 0
 DB: 3 Gaps: 0

US-10-028-384-12 (1-705) x AAC98825 (1-575)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlulysClnAspThrLeuLeuLysLeu 20
 DB 115 ATGACTAAGTTTGGATTTTGGGATTTTGGGATTTGCTATGAGAACGAGACACACTTTTGAAGCTT 174
 QY 21 LeuileLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 DB 175 CTCATTCTGTCANTGGCTGCTGATTATATCTTCTCCACTGCTGTTTCTGCTGCTGAGA 234
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 DB 235 TTTGAAAGTGTATCCATGAGTTTGATCCGCTATTATTAATTCGGACTACCAAGTTCTCTG 294
 QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspAlaTrpTyrProLeu 80

Db 295 GCTGAGGAGGGGTTTATATAATTCCTAACTGTTGATGACCGACCGCTGGTACCCCTTTG 354
 QY 81 GlyArgIleileGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100
 Db 355 GGACGAATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 414
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db 415 CATGTACTCTCAATTTTTCACATCACCATCGACATTCGGAATGTTCTGTGTCTCTGGCC 474
 QY 121 ProLeuPheSerSerPhe 126
 Db 475 CCTCTCTCTCTCTCTCTC 492

RESULT 10
 AAC00273
 ID AAC00273 standard; cDNA; 414 BP.
 XX
 AC AAC00273;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 271.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-00200610.
 XX
 PR 26-FEB-1999; 99US-0122487P.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Wilne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 DR P-PSDB; AAG00267.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX
 PS Claim 1; SEQ ID NO 271; 71pp + Sequence Listing; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors
 XX
 SQ Sequence 414 BP; 89 A; 93 C; 105 G; 127 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5,22e-96 Length: 414
 Score: 102.00 Matches: 102
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.47% Indels: 0

DB:	3	Gaps:	0
US-10-028-384-12 (1-705) x AAC00273 (1-414)			
QY	1	MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu	20
DB	108	ATGACTAGTTGGATTGCGATTGCTTATGAGAACGAGGACACACITTTGAAGCTT	167
QY	21	LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg	40
DB	168	CTCATTCGTCAATGGCTGCTATTATTCCTTCCTCACTCGTCTGTTGCTGCTCGAGA	227
QY	41	PheGluSerValLeuHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu	60
DB	228	TTTGAAGTGTATCATGAGTTTCATCGTACTTTAATTATCGGACTACCAGTTCTTG	287
QY	61	AlaGluGluGlyPheTyrLysPheHisAsnTppPheAspArgAlaTyrTyrProLeu	80
DB	288	GCTGAGGAGGGTTTATATAATCCATAACTGTTTGTATGACCGAGCCTGGTACCTTG	347
QY	81	GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr	100
DB	348	GGACGAATCATTTGGAGAACAAATTTACCAGGTTTAAATGATCACCTCTGCTGCAATCTAC	407
QY	101	HisVal	102
DB	408	CATGTA	413
RESULT 11			
ID	ABK39617	standard; cDNA; 487 BP.	
AC	ABK39617;		
DT	21-MAY-2002	(first entry)	
DE	cDNA encoding lung tumour protein clone 26496.		
KW	Lung tumour; cancer; T cell; immune response stimulator; cytostatic;		
XW	gene; ss.		
OS	Homo sapiens.		
PN	W0200204514-A2.		
PD	17-JAN-2002.		
PF	10-JUL-2001; 2001WO-US0202058.		
PR	11-JUL-2000; 2000US-00614124.		
PR	29-AUG-2000; 2000US-00651563.		
PR	08-SEP-2000; 2000US-00658824.		
PR	26-SEP-2000; 2000US-00671325.		
PR	06-OCT-2000; 2000US-00677419.		
PR	30-OCT-2000; 2000US-00702705.		
PR	13-DEC-2000; 2000US-00736457.		
PR	03-MAY-2001; 2001US-00849626.		
PA	(CORI-) CORIXA CORP.		
PI	Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;		
PI	Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS, McNabb A;		
PI	Wang A, Fanger N, Switzer A, McNeill PD, Clapper JB;		
DR	WPI; 2002-164634/21.		
XX	Novel polynucleotide encoding a lung tumor polypeptide useful for		
PT	stimulating and/or expanding T cells specific for a tumor protein.		
XX	Claim 1; SEQ ID NO 1655; 223pp; English.		
PS	The invention describes an isolated polynucleotide and polypeptide useful		
CC	for stimulating and/or expanding T cells specific for a tumour protein		
CC			

for determining the presence of a cancer in a patient. A composition containing the polynucleotide and/or polypeptide is useful for treating a lung cancer in a patient. The polypeptide is useful for removing tumour cells from a biological sample. The polynucleotide is also useful as a probe or primer to detect the level of mRNA encoding a tumour protein. This sequence encodes a lung tumour associated protein or protein fragment, described in the method of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 487 BP; 116 A; 115 C; 135 G; 121 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6,09e-96	Length:	487
Score:	102.00	Matches:	102
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	14.47%	Indels:	0
DB:	6	Gaps:	0

US-10-028-384-12 (1-705) x ABK39617 (1-487)

QY	483	TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe	502
DB	77	TACTCTTCCTCGTCCATTCCTACTATCTGCCCTGGTGGGATGGCAGTAGGATCATATTT	136
QY	503	AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysVal	522
DB	137	GATGACTTCCGAGAGCATATTTATTTGGCTCCGTCATATCTCCAGAGATCGAAGGTC	196
QY	523	MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal	542
DB	197	ATGTCCTGGTGGATTATGGCTATCAGATTACAGCTATGGCAACCCGAAACAATTTTAGG	256
QY	543	AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr	562
DB	257	GACATACACATGAGTATATACCCATATTTCTCGAGTAGGCGAGCAATGGCTCCACA	316
QY	563	GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe	582
DB	317	GAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGTATGCTGCTCATTTTT	376
QY	583	GlyGly	584
DB	377	GGAGGA	382
RESULT 12			
ID	ACA11946	standard; cDNA; 487 BP.	
AC	ACA11946;		
DT	06-JUN-2003	(first entry)	
DE	Human lung neuroendocrine carcinoma library, cDNA SEQ ID 1655.		
KW	Human; lung cancer; ss; lung tumour; cytostatic; vaccine;		
XW	T cell expansion; CD4; CD8.		
OS	Homo sapiens.		
PN	US2002197669-A1.		
PD	26-DEC-2002.		
PF	03-MAY-2001; 2001US-00849626.		
PR	13-DEC-2000; 2000US-00736457.		
PR	(BANG/) BANGUR C S.		
PA	(FANG/) FANGER G R.		
PA	(WANG/) WANG A.		

CC for treating or inhibiting development of cancer, e.g. lung cancer. This
 CC sequence represents a polynucleotide associated with the compositions and
 CC methods for the therapy and diagnosis of lung cancer

XX SQ Sequence 487 BP; 116 A; 115 C; 135 G; 121 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.09e-96 Length: 487
 Score: 102.00 Matches: 102
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.47% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-12 (1-705) x ACA03132 (1-487)

QY 483 TyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArgIleIlePhe 502
 DB 77 TACTCTTCTCCGTCATGTACTATCGCCGCTGGGGATGGCAGTAGGATCATATTT 136
 QY 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysVal 522
 DB 137 GATGACTTCGAGAGCATATTTATGCTCCGTCATATCTCCAGAGGATCGGAAGTTC 196
 QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
 DB 197 ATGCTCTGGTGGGATTATGCTATCAGATTACAGCTATGCGCAACCGAACAAATTTTAGTG 256
 QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
 DB 257 GACAATAACACATGGAATAATACCATATTTCTCGAGTAGGCGAGCAATGGCTCCACA 316
 QY 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
 DB 317 GAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGTCATATTTT 376
 QY 583 GlyGly 584
 DB 377 GGAGGA 382

RESULT 14

ID ADH47174 standard; cDNA; 487 BP.

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

XX AC ADH47174;

PT preventing and/or treating lung cancer.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX Example 1; SEQ ID NO 1655; 259pp; English.

XX 30-JUN-2000; 2000WO-US018061.
XX
XX 30-JUN-1999; 99US-00346492.
PR 15-OCT-1999; 99US-00419356.
PR 17-DEC-1999; 99US-00466867.
PR 30-DEC-1999; 99US-00476300.
PR 06-MAR-2000; 2000US-00519642.
PR 22-MAR-2000; 2000US-00519642.
PR 10-APR-2000; 2000US-00533077.
PR 27-APR-2000; 2000US-00546259.
PR 05-JUN-2000; 2000US-00560406.
XX 05-JUN-2000; 2000US-00589184.
XX

(CORI-) CORIXA CORP.

XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
PI Retter MW, Mannion J;
XX

XX WPI; 2001-071488/08.

XX Lung tumor-associated proteins and the nucleic acids that encode them,
PT useful for preventing, diagnosing and treating lung cancer.
XX

XX Claim 4; Page 282; 436pp; English.

XX The present invention describes immunogenic portions of lung tumour-
CC associated proteins (I) and the nucleic acids (NAs) that encode them. (I)
CC have cytostatic activity and can be used in gene therapy, antisense
CC inhibition and in vaccines. The NAs and the lung tumour-associated
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with their inappropriate expression,
CC especially lung cancers. For example, the NAs may be administered to
CC treat diseases by rectifying mutations or deletions in a patient's genome
CC that affect the activity of the protein by expressing inactive proteins
CC or to supplement the patients own production of (I). Additionally, the
CC NAs may be used to produce the lung-tumour associated protein, according
CC to standard recombinant DNA methodology. Conversely, antisense NA
CC molecules may be administered to down regulate protein expression by
CC binding with the cells own genes and preventing their expression. The NA
CC and complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar NA sequences in
CC samples, and hence which patients may be in need of treatment for lung
CC cancer. The (I) may be used as antigens in the production of antibodies
CC and in assays to identify modulators (agonists and antagonists) of the
CC expression and activity of the protein. AAF68083 to AAF68878 and AAF76848
CC to AAF76878 represent human lung tumour protein related nucleotide and
CC protein sequences which are used in the exemplification of the present
CC invention
XX

SQ Sequence 307 BP; 82 A; 62 C; 80 G; 83 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.3e-95	Length:	307
Score:	101.00	Matches:	101
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	14.33%	Indels:	0
DB:	5	Gaps:	0

US-10-028-384-12 (1-705) x AAF68440 (1-307)

QY	483	TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe	502
Db	3	TACTCTTCACCGTCCATTTGACTATCTGCCCGGTGGGGATGGCAGTAGGATCATATTT	62
QY	503	AspAspPheArgGluAlaTyrrTyrTrpLeuArgHisAsnThrProGluAspAlaIysVal	522
Db	63	GATGACTTCGAGAGCATATTATTGGCTCCGTCATAATACTCCAGAGGATCGAAGGTC	122
QY	523	MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal	542
Db	123	ATGTCCTGGTGGATTATTCGCTATCAGATTACAGCTATGCGAACCGAACAAATTTTAGTG	182

QY	543	AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr	562
Db	183	GACAATAACACATGGAAATAATACCCATATTTCTCGAGTAGGCAGGCAATGGCGTCCACA	242
QY	563	GluGluLysAlaIleMetArgGluLeuAspValSerTyrValLeuValIlePhe	582
Db	243	GAGGAAAAAGCCCTATGAGATCATGAGGAGCTCGATGTGAGCTATGTGCTGTCATTTT	302
QY	583	Gly	583
Db	303	GGA	305

Search completed: December 14, 2004, 23:35:26
Job time : 836.278 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 137.362 Seconds
(without alignments)
3648.073 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 705

Sequence: 1 WTKGFLSLSEKQDTLLKL.....HMLVRIYKVKDLNRLSRT 705

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 353394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10028384/runat 14122004 131646 6816/app query.fasta 1.3740
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=Oligo.rn1 -MINWATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40 cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn 1 1 259 @runat 14122004 131646 6816 -NCFU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA *

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	18.2	487	3	US-09-385-982-213
2	102	14.5	414	4	US-09-513-999C-271
3	102	14.5	487	4	US-09-702-705-1655
4	102	14.5	487	4	US-09-736-457-1655
5	102	14.5	487	4	US-09-614-124B-1655
6	102	14.5	487	4	US-09-671-325-1655
7	102	14.5	487	4	US-09-658-824-1655
8	101	14.3	307	4	US-09-702-705-375
9	101	14.3	307	4	US-09-702-705-1271
10	101	14.3	307	4	US-09-736-457-375
11	101	14.3	307	4	US-09-736-457-1271
12	101	14.3	307	4	US-09-614-124B-375

13	101	14.3	307	4	US-09-614-124B-1271	Sequence 1271, Ap
14	101	14.3	307	4	US-09-671-325-375	Sequence 375, App
15	101	14.3	307	4	US-09-671-325-1271	Sequence 1271, Ap
16	101	14.3	307	4	US-09-589-184-375	Sequence 375, App
17	101	14.3	307	4	US-09-658-824-375	Sequence 375, App
18	101	14.3	307	4	US-09-658-824-1271	Sequence 1271, Ap
19	65	9.2	307	4	US-09-702-705-588	Sequence 588, App
20	65	9.2	307	4	US-09-736-457-588	Sequence 588, App
21	65	9.2	307	4	US-09-614-124B-588	Sequence 588, App
22	65	9.2	307	4	US-09-671-325-588	Sequence 588, App
23	65	9.2	307	4	US-09-589-184-588	Sequence 588, App
24	65	9.2	307	4	US-09-658-824-588	Sequence 588, App
25	39	5.5	914	4	US-09-270-767-12856	Sequence 12856, A
26	30	4.3	1386	4	US-09-270-767-11648	Sequence 11648, A
27	26	3.7	426	4	US-09-513-999C-1113	Sequence 1213, Ap
28	26	3.7	503	4	US-09-270-767-27462	Sequence 27462, A
29	21	3.0	900	4	US-09-270-767-12331	Sequence 12331, A
30	19	2.7	291	4	US-09-313-294A-4834	Sequence 4834, Ap
31	17	2.4	268	4	US-09-313-294A-3259	Sequence 3259, Ap
32	16	2.3	2157	4	US-09-614-221A-318	Sequence 318, App
33	15	2.1	109	4	US-09-270-767-28703	Sequence 28703, A
34	15	2.1	594	4	US-09-248-796A-3090	Sequence 3090, Ap
35	13	1.8	292	4	US-09-313-294A-4442	Sequence 4442, Ap
36	13	1.8	302	4	US-09-313-294A-6869	Sequence 6869, Ap
37	12	1.7	245	4	US-09-513-999C-21090	Sequence 21090, A
38	12	1.7	299	4	US-09-313-294A-1303	Sequence 1303, Ap
39	12	1.7	507	4	US-09-513-999C-1965	Sequence 1965, Ap
40	12	1.7	560	4	US-09-270-767-1318	Sequence 1318, Ap
41	12	1.7	560	4	US-09-270-767-16600	Sequence 16600, A
42	12	1.7	616	3	US-09-328-111-332	Sequence 332, App
43	12	1.7	867	4	US-09-248-796A-3089	Sequence 3089, Ap
44	12	1.7	1660	4	US-09-270-767-158	Sequence 158, App
45	12	1.7	1660	4	US-09-270-767-158	Sequence 158, App

ALIGNMENTS

RESULT 1
US-09-385-982-213
; Sequence 213, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; PRODUCT: I I
; FILE REFERENCE: CDNA-260XX
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 213
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-385-982-213

Alignment Scores:
Pred. No.: 7.67e-121 Length: 487
Score: 128.00 Matches: 128
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 18.16% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-12 (1-705) x US-09-385-982-213 (1-487)

QY 271 PheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeu 290

Db 88 TTTGGGCTTTTGGTCTCTGCTCCAGATCCATGCTTTTGGATTACCTGCGCAGGAGTTG 147
 Qy 291 AsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPheValLeu 310
 Db 148 AATCCACAACTTTGAAGTCTTTTCCGAGGCGTCTCTCTGCTAGGCTTTTGTCTT 207
 Qy 311 LeuThrValGlyAlaLeuLeuMetLeuThrGlyLysIleSerProThrPheArgPhe 330
 Db 208 CTCACCGTGGGAGCTCTCTCATGCTGACAGGAAATATCTCCCTGGAGGGGGCTTTC 267
 Qy 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnIleProIleIleIleIleIleValSer 350
 Db 268 TACTCACTGCTGGATGCCCTCTATTGCTTAAGAACAAATCCCATCTCTCTGCTGCT 327
 Qy 351 GluHisGlnProThrThrTpsSerTyrTyrPheAspLeuGlnLeuValPheMet 370
 Db 328 GAGCATCAGCCCAACCTGGTCTCTATCTATTTTGACCTGCGAGCTCTCTGCTCATG 387
 Qy 371 PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleIle 390
 Db 388 TTTCAGTGGCTCTATTACTGCTTTAGCAACCTCTGATGCCCGGATTTTATCATC 447
 Qy 391 MetTyrGlyValThrSerMetTyr 398
 Db 448 ATGTATGCTGTGACCATGATG 471

RESULT 2

US-09-513-999C-271
 ; Sequence 271, Application US/09513999C
 ; Patent No. 6783961
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Duclert, A.
 ; APPLICANT: Giordano, J.Y.
 ; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
 ; FILE REFERENCE: 59.US2.REG
 ; CURRENT APPLICATION NUMBER: US/09/513,999C
 ; CURRENT FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/122,487
 ; PRIOR FILING DATE: 1999-02-26
 ; NUMBER OF SEQ ID NOS: 36681
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 271
 ; LENGTH: 414
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 108..413
 ; FEATURE:
 ; NAME/KEY: sig_peptide
 ; LOCATION: 108..197
 ; OTHER INFORMATION: score 7.3
 ; OTHER INFORMATION: seq LKLLLSMAAVLS/FS

US-09-513-999C-271
 Alignment Scores:
 Pred. No.: 1,81e-94 Length: 414
 Score: 102.00 Matches: 102
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.47% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-12 (1-705) x US-09-513-999C-271 (1-414)
 Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
 Db 108 ATGACTAAGTTTGGATTGCTTGGCATTTGCTATGAGAGCAGGACACACTTTTGAAGCTT 167
 Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40

Db 168 CTCACTTCTGCTCAATGGCTGCTGCTATTATTCCTTCTCCACTGCTGCTGTTGCTGCTGAGA 227
 Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 Db 228 TTTGAAGTGTATCCATGAGTTTTCATCGTACCTTAAATTAATTCGAGCTTACCAAGTTTCTG 287
 Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspAlaTTPtTyrProLeu 80
 Db 288 GCTGAGAGGGGTTTATAAATTCCTAACTGTTGATGACCGAGCTGGTACCTTTG 347
 Qy 81 GlyArgIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 348 GGCAGCAATCATTTGGAGGACAAATTTACCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 407
 Qy 101 HisVal 102
 Db 408 CATGTA 413

RESULT 3

US-09-702-705-1655
 ; Sequence 1655, Application US/09702705
 ; Patent No. 6504010
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Lodes, Michael A.
 ; APPLICANT: Fanger, Gary
 ; APPLICANT: Vedvick, Tom
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Fan, Liqun
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
 ; FILE REFERENCES: 210121.478C14
 ; CURRENT APPLICATION NUMBER: US/09/702,705
 ; CURRENT FILING DATE: 2000-10-30
 ; NUMBER OF SEQ ID NOS: 1833
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1655
 ; LENGTH: 487
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-702-705-1655

Alignment Scores:
 Pred. No.: 2.12e-94 Length: 487
 Score: 102.00 Matches: 102
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.47% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-12 (1-705) x US-09-702-705-1655 (1-487)
 Qy 483 TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
 Db 77 TACTCTTCTCCGTCCTCATTTGCTATCTCCCGTGGTGGGATGGCAGTAGATCATATTT 136
 Qy 503 AspAspPheArgGluAlaTyrTyrTyrPheuArgHisAsnThrProGluAspAlaLysVal 522
 Db 137 GATGACTTCCGAGAAGCATATTATTGGTCCGTCATATATCTCCAGAGGATGCCAAGTTC 196
 Qy 523 MetSerTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
 Db 197 ATGTCTGCTGGGATTTATGGCTATCAGATTACAGTATGGCAACCCGACAACTTTTAGTG 256
 Qy 543 AspAsnAsnThrThrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
 Db 257 GACATAACACACATGGAATAATACCCATATTTCTCGAGTAGGGCAGGCAATGGCGTCCACA 316
 Qy 563 GluGluLysAlaTyrGluIleMetArgLeuLeuAspValSerTyrValLeuValIlePhe 582

Db 317 GAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTCATTTT 376
Qy 583 GlyGly 584
Db 377 GGAGGA 382

RESULT 4

US-09-736-457-1655
; Sequence 1655, Application US/09736457
; Patent No. 6509448

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedrick, Tom
; APPLICANT: Carter, Darriek
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.478C15

; CURRENT APPLICATION NUMBER: US/09/736,457

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 1864

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1655

; LENGTH: 487

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-736-457-1655

Alignment Scores:

Pred. No.: 2,12e-94 Length: 487
Score: 102.00 Matches: 102
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.47% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-736-457-1655 (1-487)

Qy 483 TyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArgIleIlePhe 502
Db 77 TACTCTTCCTCGTCCATGTAATATCCCATATTTCTCGAGTAGGAGCGCATGCGTCCACA 316
Qy 503 AspAspPheArgGluAlaTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
Db 137 GATGACTTCGAGAGCATATTTATGGCTCCGTCAATAATCTCCAGAGGATCGAAGGTC 196

Qy 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
Db 197 ATGCTCTGGGGATTATGGCTATCAGATTACAGCTATGCGCAACCAATTTTAGTG 256

Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 257 GACAATAACACATGGAATATATCCCATATTTCTCGAGTAGGAGCGCATGCGTCCACA 316

Qy 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 317 GAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTCATTTT 376

Qy 583 GlyGly 584

Db 377 GGAGGA 382

RESULT 5

US-09-614-124B-1655

; Sequence 1655, Application US/09614124B

; Patent No. 6630574

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vedrick, Tom

; APPLICANT: Carter, Darriek

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; FILE REFERENCE: 210121.478C9

; CURRENT APPLICATION NUMBER: US/09/614,124B

; CURRENT FILING DATE: 2001-07-11

; NUMBER OF SEQ ID NOS: 1668

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1655

; LENGTH: 487

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-614-124B-1655

Alignment Scores:

Pred. No.: 2,12e-94 Length: 487
Score: 102.00 Matches: 102
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.47% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-614-124B-1655 (1-487)

Qy 483 TyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArgIleIlePhe 502
Db 77 TACTCTTCCTCGTCCATGTAATATCCCATATTTCTCGAGTAGGAGCGCATATTT 136

Qy 503 AspAspPheArgGluAlaTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
Db 137 GATGACTTCGAGAGCATATTTATGGCTCCGTCAATAATCTCCAGAGGATCGAAGGTC 196

Qy 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
Db 197 ATGCTCTGGGGATTATGGCTATCAGATTACAGCTATGCGCAACCAATTTTAGTG 256

Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 257 GACAATAACACATGGAATATATCCCATATTTCTCGAGTAGGAGCGCAATGGCGTCCACA 316

Qy 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 317 GAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTCATTTT 376

Qy 583 GlyGly 584

Db 377 GGAGGA 382

RESULT 6

US-09-671-325-1655

; Sequence 1655, Application US/09671325

; Patent No. 6867154

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vedrick, Tom

; APPLICANT: Carter, Darriek

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.478C12

487

QY 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaIysVal 522
Db 63 GATGACTTCGAGAGCATAATTATGGCTCCGTCTAATAATCTCCAGAGATGCGAAGGTC 122
QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
Db 123 ATGTCCTGGTGGGATTATGGCTATCAGATTACAGCTATGCGCAACCGAACAAATTTAGTG 182
QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 183 GACATAACACATGGAATAATACCCATAATTTCTCAGTAGGCGAGGCAATGGCGTCCACA 242
QY 563 GluGluLysAlaTyrGluIleValMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 243 GAGGAAAAAGCCTATGAGATCATGAGGAGGCTCGATGTCAGCTAIGTCTGTCATTTT 302
QY 583 Gly 583
Db 303 GGA 305

RESULT 9

US-09-702-705-1271
; Sequence 1271, Application US/09702705
; Patent No. 6504010

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14

; CURRENT APPLICATION NUMBER: US/09/702,705

; CURRENT FILING DATE: 2000-10-30

; NUMBER OF SEQ ID NOS: 1833

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1271

; LENGTH: 307

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-702-705-1271

Alignment Scores:

Pred. No.: 1.42e-93 Length: 307
Score: 101.00 Matches: 101
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.33% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-702-705-1271 (1-307)

QY 483 TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
Db 3 TACTCTTCCTCCGTCATATGCTATCTGCGCGTGGTGGGATGCGAGTATCATATTT 62
QY 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaIysVal 522
Db 63 GATGACTTCGAGAGCATAATTATGGCTCCGTCTAATAATCTCCAGAGATGCGAAGGTC 122
QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
Db 123 ATGTCCTGGTGGGATTATGGCTATCAGATTACAGCTATGCGCAACCGAACAAATTTAGTG 182
QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 183 GACATAACACATGGAATAATACCCATAATTTCTCAGTAGGCGAGGCAATGGCGTCCACA 242

QY 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 243 GAGGAAAAAGCCTATGAGATCATGAGGAGGCTCGATGTCAGCTATGCTGTCATTTT 302
QY 583 Gly 583
Db 303 GGA 305

RESULT 10

US-09-736-457-375
; Sequence 375, Application US/09736457
; Patent No. 6509448

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15

; CURRENT APPLICATION NUMBER: US/09/736,457

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 1864

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 375

; LENGTH: 307

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-736-457-375

Alignment Scores:

Pred. No.: 1.42e-93 Length: 307
Score: 101.00 Matches: 101
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.33% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-736-457-375 (1-307)

QY 483 TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
Db 3 TACTCTTCCTCCGTCATATGCTATCTGCGCGTGGTGGGATGCGAGTATCATATTT 62
QY 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaIysVal 522
Db 63 GATGACTTCGAGAGCATAATTATGGCTCCGTCTAATAATCTCCAGAGATGCGAAGGTC 122
QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
Db 123 ATGTCCTGGTGGGATTATGGCTATCAGATTACAGCTATGCGCAACCGAACAAATTTAGTG 182
QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 183 GACATAACACATGGAATAATACCCATAATTTCTCAGTAGGCGAGGCAATGGCGTCCACA 242
QY 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 243 GAGGAAAAAGCCTATGAGATCATGAGGAGGCTCGATGTCAGCTATGCTGTCATTTT 302
QY 583 Gly 583
Db 303 GGA 305

RESULT 11

US-09-736-457-1271
; Sequence 1271, Application US/09736457

```

; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Pan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1271
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-736-457-1271

Alignment Scores:
Pred. No.: 1.42e-93 Length: 307
Score: 101.00 Matches: 101
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.33% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-736-457-1271 (1-307)

Qy 483 TyrSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
Db 3 TACTCTTCCTCGTCATTTACTATCTGCCCCGTGGGGATGGCAGTAGGATCATATTT 62
Qy 503 AspAspPheArgGluAlaTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
Db 63 GATGACTTCGAGAAGCATATTTATGGCTCCGCTCATATACTCCAGAGATGGCAAGTC 122
Qy 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleVal 542
Db 123 ATGCTCGTGGGATTTAGCTATCAGATTACAGTATGGCAACCGAACAATTTAGTG 182
Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 183 GACAATAACACATGGATAATACCATATTTCTCGAGTAGGCGAGCAATGGCTCCACA 242
Qy 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 243 GAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTCATTTT 302
Qy 583 Gly 583
Db 303 GGA 305

RESULT 12
US-09-614-124B-375
; Sequence 375, Application US/09614124B
; Patent No. 6630574
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1271
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-1271

Alignment Scores:

```

```

; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 375
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-375

Alignment Scores:
Pred. No.: 1.42e-93 Length: 307
Score: 101.00 Matches: 101
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.33% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-614-124B-375 (1-307)

Qy 483 TyrSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
Db 3 TACTCTTCCTCGTCATTTACTATCTGCCCCGTGGGGATGGCAGTAGGATCATATTT 62
Qy 503 AspAspPheArgGluAlaTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
Db 63 GATGACTTCGAGAAGCATATTTATGGCTCCGCTCATATACTCCAGAGATGGCAAGTC 122
Qy 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleVal 542
Db 123 ATGCTCGTGGGATTTAGCTATCAGATTACAGTATGGCAACCGAACAATTTAGTG 182
Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 183 GACAATAACACATGGATAATACCATATTTCTCGAGTAGGCGAGCAATGGCTCCACA 242
Qy 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 243 GAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTCATTTT 302
Qy 583 Gly 583
Db 303 GGA 305

RESULT 13
US-09-614-124B-1271
; Sequence 1271, Application US/09614124B
; Patent No. 6630574
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1271
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-1271

Alignment Scores:

```

Qy	563	GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe	582
Db	243	GAGGAAAGCCCTAAGAGATCATGAGGAGCTCGATGTCAGCTATGTCGTGTCATTTT	302
Qy	583	Gly	583
Db	303	GGA	305

Search completed: December 14, 2004, 23:43:59
 Job time : 141.362 secs

Pred. No.: 1.42e-93 Length: 307
 Score: 101.00 Matches: 101
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.33% Indels: 0
 DB: Gaps: 0

US-10-028-384-12 (1-705) x US-09-614-124B-1271 (1-307)

Qy 483 TyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArgIleIlePhe 502
 Db 3 TACTCTTCCTCGGATGATGCTATCGCGGTGGGATGGCAGTAGGATCATATTT 62
 Qy 503 AspAspPheArgGluAlaTyTrpLeuArgHisAsnThrProGluAspAlaLysVal 522
 Db 63 GATGACTTCGGAAGCATATTAATGGCTCGGATGATGCTATCGCGGTGGGATGGCAGTAGGATCATATTT 62
 Qy 523 MetSerTrpAspTyTrpGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
 Db 123 ATGCTCTGGTGGATGATGCTATCGATTCAGATTACAGCTATGCAACCGAACATTTTAGTG 182
 Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
 Db 183 GACAATAACATCGAATAATACCCATATTTCTCGAGTAGGGCAGCAATGGCGTCCACA 242
 Qy 563 GluGluLysAlaTyTrpGlnIleMetArgGluLeuAspValSerTyValLeuValIlePhe 582
 Db 243 GAGGAAAAGCCATGATGATCATGAGGAGCTCGATGTCAGCTATGCTGGTCAATTTT 302
 Qy 583 Gly 583
 Db 303 GGA 305

RESULT 14

US-09-671-325-375
 ; Sequence 375, Application US/09671325
 ; Patent No. 6667154

; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vecvick, Tom

; APPLICANT: Carter, Darrick

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.478C12

; CURRENT APPLICATION NUMBER: US/09/671,325

; CURRENT FILING DATE: 2000-09-26

; NUMBER OF SEQ ID NOS: 1825

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 375

; LENGTH: 307

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-671-325-375

Alignment Scores:

Pred. No.: 1.42e-93 Length: 307
 Score: 101.00 Matches: 101
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.33% Indels: 0
 DB: Gaps: 0

US-10-028-384-12 (1-705) x US-09-671-325-375 (1-307)

Qy 483 TyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArgIleIlePhe 502
 Db 3 TACTCTTCCTCGGATGATGCTATCGCGGTGGGATGGCAGTAGGATCATATTT 62

Qy 503 AspAspPheArgGluAlaTyTrpLeuArgHisAsnThrProGluAspAlaLysVal 522
 Db 63 GATGACTTCGGAAGCATATTAATGGCTCGGATGATGCTATCGCGGTGGGATGGCAGTAGGATCATATTT 62
 Qy 523 MetSerTrpAspTyTrpGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
 Db 123 ATGCTCTGGTGGATGATGCTATCGATTCAGATTACAGCTATGCAACCGAACATTTTAGTG 182
 Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
 Db 183 GACAATAACATCGAATAATACCCATATTTCTCGAGTAGGGCAGCAATGGCGTCCACA 242
 Qy 563 GluGluLysAlaTyTrpGlnIleMetArgGluLeuAspValSerTyValLeuValIlePhe 582
 Db 243 GAGGAAAAGCCATGATGATCATGAGGAGCTCGATGTCAGCTATGCTGGTCAATTTT 302
 Qy 583 Gly 583
 Db 303 GGA 305

RESULT 15

US-09-671-325-1271

; Sequence 1271, Application US/09671325

; Patent No. 6667154

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vecvick, Tom

; APPLICANT: Carter, Darrick

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.478C12

; CURRENT APPLICATION NUMBER: US/09/671,325

; CURRENT FILING DATE: 2000-09-26

; NUMBER OF SEQ ID NOS: 1825

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1271

; LENGTH: 307

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-671-325-1271

Alignment Scores:

Pred. No.: 1.42e-93 Length: 307
 Score: 101.00 Matches: 101
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.33% Indels: 0
 DB: Gaps: 0

US-10-028-384-12 (1-705) x US-09-671-325-1271 (1-307)

Qy 483 TyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArgIleIlePhe 502
 Db 3 TACTCTTCCTCGGATGATGCTATCGCGGTGGGATGGCAGTAGGATCATATTT 62
 Qy 503 AspAspPheArgGluAlaTyTrpLeuArgHisAsnThrProGluAspAlaLysVal 522
 Db 63 GATGACTTCGGAAGCATATTAATGGCTCGGATGATGCTATCGCGGTGGGATGGCAGTAGGATCATATTT 62
 Qy 523 MetSerTrpAspTyTrpGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
 Db 123 ATGCTCTGGTGGATGATGCTATCGATTCAGATTACAGCTATGCAACCGAACATTTTAGTG 182
 Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
 Db 183 GACAATAACATCGAATAATACCCATATTTCTCGAGTAGGGCAGCAATGGCGTCCACA 242